

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11) EP 0 786 519 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
30.07.1997 Bulletin 1997/31

(51) Int Cl. 6: C12N 15/00

(21) Application number: 97100117.7

(22) Date of filing: 07.01.1997

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE
Designated Extension States:
AL LT LV RO SI

(30) Priority: 05.01.1996 US 9861

(71) Applicant: HUMAN GENOME SCIENCES, INC.
Rockville, MD 20850-3338 (US)

(72) Inventors:
• Kunsch, Charles A.
Gaithersburg, Maryland 20882 (US)

- Chol, Gil H.
Rockville, Maryland 20850 (US)
- Barash, Steven C.
Rockville, Maryland 20850 (US)
- Dillon, Patrick J.
Gaithersburg, Maryland 20879 (US)
- Fannon, Michael R.
Silver Spring, Maryland 20906 (US)
- Rosen, Craig A.
Laytonsville, Maryland 20882 (US)

(74) Representative: VOSSIUS & PARTNER
Postfach 86 07 67
81634 München (DE)

(54) Staphylococcus aureus polynucleotides and sequences

(57) The present invention provides polynucleotide sequences of the genome of *Staphylococcus aureus*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynu-

cleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

EP 0 786 519 A2

Description

The present invention relates to the field of molecular biology. In particular, it relates to, among other things, nucleotide sequences of *Staphylococcus aureus*, contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequences, and uses of the polynucleotides and sequences thereof, such as in fermentation, polypeptide production, assays and pharmaceutical development, among others.

5 The genus *Staphylococcus* includes at least 20 distinct species. (For a review see Novick, R. P., *The Staphylococcus as a Molecular Genetic System*, Chapter 1, pgs. 1-37 in MOLECULAR BIOLOGY OF THE STAPHYLOCOCCI

R. Novick, Ed., VCH Publishers, New York (1990)). Species differ from one another by 80% or more, by hybridization

10 kinetics, whereas strains within a species are at least 90% identical by the same measure.

The species *Staphylococcus aureus*, a gram-positive, facultatively aerobic, clump-forming cocci, is among the most important etiological agents of bacterial infection in humans, as discussed briefly below.

Human Health and *S. Aureus*

15 *Staphylococcus aureus* is a ubiquitous pathogen. (See, for instance, Mims et al., MEDICAL MICROBIOLOGY, Mosby-Year Book Europe Limited, London, UK (1993)). It is an etiological agent of a variety of conditions, ranging in severity from mild to fatal. A few of the more common conditions caused by *S. aureus* infection are burns, cellulitis, 20 eyelid infections, food poisoning, joint infections, neonatal conjunctivitis, osteomyelitis, skin infections, surgical wound infection, scalded skin syndrome and toxic shock syndrome, some of which are described further below.

Burns

25 Burn wounds generally are sterile initially. However, they generally compromise physical and immune barriers to infection, cause loss of fluid and electrolytes and result in local or general physiological dysfunction. After cooling, contact with viable bacteria results in mixed colonization at the injury site. Infection may be restricted to the non-viable debris on the burn surface ("eschar"), it may progress into full skin infection and invade viable tissue below the eschar and it may reach below the skin, enter the lymphatic and blood circulation and develop into septicaemia. *S. aureus* is 30 among the most important pathogens typically found in burn wound infections. It can destroy granulation tissue and produce severe septicaemia.

Cellulitis

35 Cellulitis, an acute infection of the skin that expands from a typically superficial origin to spread below the cutaneous layer, most commonly is caused by *S. aureus* in conjunction with *S. pyogenes*. Cellulitis can lead to systemic infection. In fact, cellulitis can be one aspect of synergistic bacterial gangrene. This condition typically is caused by a mixture of *S. aureus* and microaerophilic streptococci. It causes necrosis and treatment is limited to excision of the necrotic tissue. The condition often is fatal.

40 Eyelid infections

S. aureus is the cause of styes and of "sticky eye" in neonates, among other eye infections. Typically such infections are limited to the surface of the eye, and may occasionally penetrate the surface with more severe consequences.

45 Food poisoning

Some strains of *S. aureus* produce one or more of five serologically distinct, heat and acid stable enterotoxins that are not destroyed by digestive process of the stomach and small intestine (enterotoxins A-E). Ingestion of the toxin, in sufficient quantities, typically results in severe vomiting, but not diarrhoea. The effect does not require viable bacteria.

50 Although the toxins are known, their mechanism of action is not understood.

Joint infections

55 *S. aureus* infects bone joints causing diseases such osteomyelitis.

Osteomyelitis

S. aureus is the most common causative agent of haematogenous osteomyelitis. The disease tends to occur in

children and adolescents more than adults and it is associated with non-penetrating injuries to bones. Infection typically occurs in the long end of growing bone, hence its occurrence in physically immature populations. Most often, infection is localized in the vicinity of sprouting capillary loops adjacent to epiphyseal growth plates in the end of long, growing bones.

5

Skin infections

S. aureus is the most common pathogen of such minor skin infections as abscesses and boils. Such infections often are resolved by normal host response mechanisms, but they also can develop into severe internal infections.

10

Recurrent infections of the nasal passages plague nasal carriers of *S. aureus*.

Surgical Wound Infections

Surgical wounds often penetrate far into the body. Infection of such wound thus poses a grave risk to the patient.

15

S. aureus is the most important causative agent of infections in surgical wounds. *S. aureus* is unusually adept at invading surgical wounds; sutured wounds can be infected by far fewer *S. aureus* cells than are necessary to cause infection in normal skin. Invasion of surgical wound can lead to severe *S. aureus* septicaemia. Invasion of the blood stream by *S. aureus* can lead to seeding and infection of internal organs, particularly heart valves and bone, causing systemic diseases, such as endocarditis and osteomyelitis.

20

Scalded Skin Syndrome

S. aureus is responsible for "scalded skin syndrome" (also called toxic epidermal necrolysis, Ritter's disease and Lyell's disease). This disease occurs in older children, typically in outbreaks caused by flowering of *S. aureus* strains 25 produce exfoliation(also called scalded skin syndrome toxin). Although the bacteria initially may infect only a minor lesion, the toxin destroys intercellular connections, spreads epidermal layers and allows the infection to penetrate the outer layer of the skin, producing the desquamation that typifies the disease. Shedding of the outer layer of skin generally reveals normal skin below, but fluid lost in the process can produce severe injury in young children if it is not treated properly.

30

Toxic Shock Syndrome

Toxic shock syndrome is caused by strains of *S. aureus* that produce the so-called toxic shock syndrome toxin. The disease can be caused by *S. aureus* infection at any site, but it is too often erroneously viewed exclusively as a disease solely of women who use tampons. The disease involves toxæmia and septicaemia, and can be fatal.

Nocosomial Infections

In the 1984 National Nocosomial Infection Surveillance Study ("NNIS") *S. aureus* was the most prevalent agent 40 of surgical wound infections in many hospital services, including medicine, surgery, obstetrics, pediatrics and newborns.

*Resistance to drugs of *S. aureus* strains*

Prior to the introduction of penicillin the prognosis for patients seriously infected with *S. aureus* was unfavorable.

45

Following the introduction of penicillin in the early 1940s even the worst *S. aureus* infections generally could be treated successfully. The emergence of penicillin-resistant strains of *S. aureus* did not take long, however. Most strains of *S. aureus* encountered in hospital infections today do not respond to penicillin; although, fortunately, this is not the case for *S. aureus* encountered in community infections.

It is well known now that penicillin-resistant strains of *S. aureus* produce a lactamase which converts penicillin to 50 penicillinoic acid, and thereby destroys antibiotic activity. Furthermore, the lactamase gene often is propagated episomally, typically on a plasmid, and often is only one of several genes on an episomal element that, together, confer multidrug resistance.

Methicillins, introduced in the 1960s, largely overcame the problem of penicillin resistance in *S. aureus*. These 55 compounds conserve the portions of penicillin responsible for antibiotic activity and modify or alter other portions that make penicillin a good substrate for inactivating lactamases. However, methicillin resistance has emerged in *S. aureus*, along with resistance to many other antibiotics effective against this organism, including aminoglycosides, tetracycline, chloramphenicol, macrolides and lincosamides. In fact, methicillin-resistant strains of *S. aureus* generally are multiply drug resistant.

The molecular genetics of most types of drug resistance in *S. aureus* has been elucidated (See Lyon *et al.*, *Microbiology Reviews* 51: 88-134 (1987)). Generally, resistance is mediated by plasmids, as noted above regarding penicillin resistance; however, several stable forms of drug resistance have been observed that apparently involve integration of a resistance element into the *S. aureus* genome itself.

5 Thus far each new antibiotic gives rise to resistance strains, stains emerge that are resistance to multiple drugs and increasingly persistent forms of resistance begin to emerge. Drug resistance of *S. aureus* infections already poses significant treatment difficulties, which are likely to get much worse unless new therapeutic agents are developed.

10 Molecular Genetics of *Staphylococcus Aureus*

Despite its importance in, among other things, human disease, relatively little is known about the genome of this organism.

15 Most genetic studies of *S. aureus* have been carried out using the the strain NCTC8325, which contains prophages psi11 psi12 and psi13, and the UV-cured derivative of this strain, 8325-4 (also referred to as RN450), which is free of the prophages.

These studies revealed that the *S. aureus* genome, like that of other *staphylococci*, consists of one circular, covalently closed, double-stranded DNA and a collection of so-called variable accessory genetic elements, such as prophages, plasmids, transposons and the like.

20 Physical characterization of the genome has not been carried out in any detail. Pattee *et al.* published a low resolution and incomplete genetic and physical map of the chromosome of *S. aureus* strain NCTC 8325. (Pattee *et al.* Genetic and Physical Mapping of Chromosome of *Staphylococcus aureus* NCTC 8325, Chapter 11, pgs. 163-169 in. MOLECULAR BIOLOGY OF THE STAPHYLOCOCCI, R.P. Novick, Ed., VCH Publishers, New York, (1990) The genetic map largely was produced by mapping insertions of Tn551 and Tn4001, which, respectively, confer erythromycin and gentamicin resistance, and by analysis of SmaI-digested DNA by Pulsed Field Gel Electrophoresis ("PFGE").

25 The map was of low resolution; even estimating the physical size of the genome was difficult, according to the investigators. The size of the largest SmaI chromosome fragment, for instance, was too large for accurate sizing by PFGE. To estimate its size, additional restriction sites had to be introduced into the chromosome using a transposon containing a SmaI recognition sequence.

30 In sum, most physical characteristics and almost all of the genes of *Staphylococcus aureus* are unknown. Among the few genes that have been identified, most have not been physically mapped or characterized in detail. Only a very few genes of this organism have been sequenced. (See, for instance Thornsberry, J., *Antimicrobial Chemotherapy* 21 Suppl C: 9-16 (1988), current versions of GENBANK and other nucleic acid databases, and references that relate to the genome of *S. aureus* such as those set out elsewhere herein.)

35 It is clear that the etiology of diseases mediated or exacerbated by *S. aureus* infection involves the programmed expression of *S. aureus* genes, and that characterizing the genes and their patterns of expression would add dramatically to our understanding of the organism and its host interactions. Knowledge of *S. aureus* genes and genomic organization would dramatically improve understanding of disease etiology and lead to improved and new ways of preventing, ameliorating, arresting and reversing diseases. Moreover, characterized genes and genomic fragments of *S. aureus* would provide reagents for, among other things, detecting, characterizing and controlling *S. aureus* infections.

40 There is a need therefore to characterize the genome of *S. aureus* and for polynucleotides and sequences of this organism.

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS: 1-5,191.

45 The present invention provides the nucleotide sequence of several thousand contigs of the *Staphylococcus aureus* genome, which are listed in tables below and set out in the Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS:1-5,191.

50 The present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequences of SEQ ID NOS:1-5,191.

The nucleotide sequence of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence which is at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequence of SEQ ID NOS:1-5,191 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to:magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

The present invention further provides systems, particularly computer-based systems which contain the sequence

information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Staphylococcus aureus* genome.

Another embodiment of the present invention is directed to fragments, preferably isolated fragments, of the *Staphylococcus aureus* genome having particular structural or functional attributes. Such fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs, " fragments which modulate the expression of an operably linked ORF, hereinafter referred to as expression modulating fragments or EMFs," and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter referred to as diagnostic fragments or "DFs."

Each of the ORFs in fragments of the *Staphylococcus aureus* genome disclosed in Tables 1-3, and the EMFs found 5' to the ORFs, can be used in numerous ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

The present invention further includes recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genome of the present invention. The recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Staphylococcus aureus* genome of the present invention. The host cells can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a prokaryotic cell such as a bacterial cell.

The present invention is further directed to polypeptides and proteins, preferably isolated polypeptides and proteins, encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present invention can from cells which have been altered to express them.

The invention further provides polypeptides, preferably isolated polypeptides, comprising *Staphylococcus aureus* epitopes and vaccine compositions comprising such polypeptides. Also provided are methods for vaccinating an individual against *Staphylococcus aureus* infection.

The invention further provides methods of obtaining homologs of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both monoclonal and polyclonal antibodies.

The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the antibodies of the present invention, or one or more of the DFs or antigens of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the antibodies, antigens, or one of the DFs of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of bound antibodies, antigens or hybridized DFs.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of: (a)contacting an agent with an isolated protein encoded by one of the ORFs of the present invention; and (b)determining whether the agent binds to said protein.

The present genomic sequences of *Staphylococcus aureus* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Staphylococcus aureus* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to *Staphylococcus aureus* researchers and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes

has and will greatly enhance the ability to analyze and understand chromosomal organization. In particular, sequenced contigs and genomes will provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the ability to do comparative genomic and molecular phylogeny.

5 FIGURE 1 is a block diagram of a computer system (102) that can be used to implement computer-based systems of present invention.

10 FIGURE 2 is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Staphylococcus aureus* genome of the present invention. Both Macintosh and Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of sequence files. The program Loadis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Staphylococcus aureus* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using extrseq, a Unix utility for retrieving sequences from an SQL database. The resulting sequence file is processed by seq_filter to trim portions of the sequences with more than 2% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research (TIGR*) for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly step is loaded into the database with the lassie program. Identification of open reading frames (ORFs) is accomplished by processing contigs with zorf. The ORFs are searched against *S. aureus* sequences from Genbank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990). Results of the ORF determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3.

15 25 The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS:1-5,191. (As used herein, the "primary sequence" refers to the nucleotide sequence represented by the IUPAC nomenclature system.)

20 30 In addition to the aforementioned *Staphylococcus aureus* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS:1-5,191, or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

35 40 As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS:1-5,191" refers to any portion of the SEQ ID NOS:1-5,191 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Staphylococcus aureus* open reading frames (ORFs), expression modulating fragment (EMFs*) and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in sample ("DFs"). A non-limiting identification of preferred representative fragments is provided in Tables 1-3.

45 50 As discussed in detail below, the information provided in SEQ ID NOS:1-5,191 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled in the art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Staphylococcus aureus* proteins.

55 60 While the presently disclosed sequences of SEQ ID NOS:1-5,191 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances, further investigation of a fragment or sequence of the invention may reveal a nucleotide sequence error present in a nucleotide sequence disclosed in SEQ ID NOS:1-5,191. However, once the present invention is made available (*i.e.*, once the information in SEQ ID NOS:1-5,191 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS:1-5,191 will be well within the skill of the art. The present disclosure makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotide may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

65 70 Even if all of the very rare sequencing errors in SEQ ID NOS:1-5,191 were corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191.

75 80 As discussed elsewhere hererin, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning and sequencing DNA. Detailed methods for obtaining

libraries and for sequencing are provided below, for instance. A wide variety of *Staphylococcus aureus* strains that can be used to prepare *S. aureus* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC*).

5 The nucleotide sequences of the genomes from different strains of *Staphylococcus aureus* differ somewhat. However, the nucleotide sequences of the genomes of all *Staphylococcus aureus* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS:1-5,191. Nearly all will be at least 99% identical and the great majority will be 99.9% identical.

10 Thus, the present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191, in a form which can be readily used, analyzed and interpreted by the skilled artisan.

15 Methods for determining whether a nucleotide sequence is at least 95%, at least 99% or at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191 are routine and readily available to the skilled artisan. For example, the well known fasta algorithm described in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85: 2444 (1988) can be used to generate the percent identity of nucleotide sequences. The BLASTN program also can be used to generate an identity score of polynucleotides compared to one another.

COMPUTER RELATED EMBODIMENTS

20 The nucleotide sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-5,191 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; *i.e.*, a nucleotide sequence provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide of SEQ ID NOS:1-5,191. Such a manufacture provides a large portion of the *Staphylococcus aureus* genome and parts thereof (*e.g.*, a *Staphylococcus aureus* open reading frame (ORF)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the *Staphylococcus aureus* genome or a subset thereof as it exists in nature or in purified form.

25 In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD- ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

30 As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

35 A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

40 Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS:1-5,191 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

45 The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol.* 215:403410 (1990)) and BLAZE (Brutlag *et al.*, *Comp. Chem.* 17:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Staphylococcus aureus* genome which contain homology to ORFs or proteins from both *Staphylococcus aureus* and from other organisms. Among the ORFs discussed

herein are protein encoding fragments of the *Staphylococcus aureus* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, commercially important fragments of the *Staphylococcus aureus* genome.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Staphylococcus aureus* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the *Staphylococcus aureus* genome. In the present examples, implementing software which implement the BLAST and BLAZE algorithms, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990), was used to identify open reading frames within the *Staphylococcus aureus* genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, etc. A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, etc.) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for

accessing and processing the genomic sequence (such as search tools, comparing tools, etc.) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

5 **BIOCHEMICAL EMBODIMENTS**

Other embodiments of the present invention are directed to fragments of the *Staphylococcus aureus* genome, preferably to isolated fragments. The fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to fragments which encode peptides, hereinafter open reading frames (ORFs), fragments which 10 modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter diagnostic fragments (DFs).

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Staphylococcus aureus* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS:1-5,191, to 15 representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

A variety of purification means can be used to generate the isolated fragments of the present invention. These 20 include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Staphylococcus aureus* DNA can be mechanically sheared to produce fragments of 15-20 kb in length. These fragments can then be used to generate an *Staphylococcus aureus* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated 25 in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS: 1-5,191. Well known and routine techniques of PCR cloning then can be used to isolate the ORF from the lambda DNA library of *Staphylococcus aureus* genomic DNA. Thus, given the availability of SEQ ID NOS:1-5,191, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS:1-5,191 using methods set out above, those of skill will be enabled by the present disclosure to isolate any ORF-containing or other nucleic acid fragment of the present invention.

30 The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA.

As used herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any 35 termination codons and is a sequence translatable into protein.

Tables 1, 2 and 3 list ORFs in the *Staphylococcus aureus* genomic contigs of the present invention that were 40 identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It will be appreciated that other criteria can be used, in accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive or more selective lists.

Table 1 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are at least 80 amino acids long and over a continuous region of at least 50 bases which are 95% or more identical (by BLAST analysis) to 45 an *S. aureus* nucleotide sequence available through Genbank in November 1996.

Table 2 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through Genbank by September 1996.

Table 3 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that do not match significantly, 50 by BLASTP analysis, a polypeptide sequence available through Genbank by September 1996.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column indicates the reading frame, taking the first 5' nucleotide of the contig as the start of the +1 frame; the fourth column indicates the first nucleotide of the ORF, counting from the 5' end of the contig strand; and the fifth column indicates the length of each ORF in nucleotides.

55 In Tables 1 and 2, column six, lists the Reference* for the closest matching sequence available through Genbank. These reference numbers are the databases entry numbers commonly used by those of skill in the art, who will be familiar with their denominators. Descriptions of the nomenclature are available from the National Center for Biotechnology Information. Column seven in Tables 1 and 2 provides the gene name* of the matching sequence; column eight provides the BLAST identity* score from the comparison of the ORF and the homologous gene; and column nine indicates the length in nucleotides of the highest scoring segment pair* identified by the BLAST identity analysis.

In Table 3, the last column, column six, indicates the length of each ORF in amino acid residues.

The concepts of percent identity and percent similarity of two polypeptide sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (e.g., at positions

1, 3 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (i.e., possessed similar biochemical characteristics). Many programs for analysis of nucleotide or amino acid sequence similarity, such as fasta and BLAST specifically list per cent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the per cent identity* of the highest scoring segment pair* in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in the pertinent literature highlighted by the citations provided below.

5 It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Staphylococcus aureus* genome other than those listed in Tables 10 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

15 As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF.

15 As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are fragments which induce the expression or an operably linked ORF in response to a specific regulatory factor or physiological event.

20 EMF sequences can be identified within the contigs of the *Staphylococcus aureus* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the *Staphylococcus aureus* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a target sequence or target motif in the computer-based systems 25 of the present invention. Further, the two methods can be combined and used together.

25 The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic 30 resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below.

30 A sequence which is suspected as being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host is examined under appropriate conditions.

35 As described above, an EMF will modulate the expression of an operably linked marker sequence.

35 As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Staphylococcus aureus* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Staphylococcus aureus* genome, such as by using well-known computer analysis software, and by generating and 40 testing probes or amplification primers consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

40 The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS:1-5, 191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably 99% and most preferably 99.9% identical to SEQ ID NOS:1-5, 191, with a sequence 45 from another isolate of the same species.

45 Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the nucleic acid sequences mentioned above. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated.

50 Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (i.e., sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Staphylococcus aureus* origin isolated by using part or all of the fragments 55 in question as a probe or primer.

55 Each of the ORFs of the *Staphylococcus aureus* genome disclosed in Tables 1, 2 and 3, and the EMFs found to the ORFs, can be used as polynucleotide reagents in numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particularly *Staphylococcus aureus*. Especially preferred in this regard are ORF such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Staphylococcus aureus*. Also particularly preferred are ORFs that can be used to distinguish between strains of *Staphylococcus aureus*.

staphylococcus aureus, particularly those that distinguish medically important strain, such as drug-resistant strains.

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation, or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for example, Lee *et al.*, *Nucl. Acids Res.* 6: 3073 (1979); Cooney *et al.*, *Science* 241: 456 (1988); and Dervan *et al.*, *Science* 251: 1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, *J. Neurochem.* 56: 560 (1991) and OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press, Boca Raton, FL (1988).

The present invention further provides recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBluescript SK and KS (+ and -), pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (available from Stratagene), pSVK3, pBPV, pMSG, pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein- I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

The present invention further provides host cells containing any one of the isolated fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or a prokaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE, dextran mediated transfection and electroporation, which are described in, for instance, Davis, L. *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986).

A host cell containing one of the fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (e.g., an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs depicted in Tables 2 and 3 which encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Such short fragments as may be obtained most readily by synthesis are useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polypeptides and proteins to isolate and purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immu-

no-affinity chromatography.

The polypeptides and proteins of the present invention also can be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and SF9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (e.g., microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, e.g., *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Staphylococcus aureus* genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

"Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancers and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), alpha-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *Staphylococcus aureus*, *E. coli*, *B. subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*. Others

may, also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available from Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, *Cell* 23: 175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

An additional aspect of the invention includes *Staphylococcus aureus* polypeptides which are useful as immuno-diagnostic antigens and/or immunoprotective vaccines, collectively "immunologically useful polypeptides". Such immunologically useful polypeptides may be selected from the ORFs disclosed herein based on techniques well known in the art and described elsewhere herein. The inventors have used the following criteria to select several immunologically useful polypeptides:

As is known in the art, an amino terminal type I signal sequence directs a nascent protein across the plasma and outer membranes to the exterior of the bacterial cell. Such outermembrane polypeptides are expected to be immunologically useful. According to Izard, J. W. et al., *Mol. Microbiol.* 13, 765-773; (1994), polypeptides containing type I signal sequences contain the following physical attributes: The length of the type I signal sequence is approximately 15 to 25 primarily hydrophobic amino acid residues with a net positive charge in the extreme amino terminus; the central region of the signal sequence must adopt an alpha-helical conformation in a hydrophobic environment; and the region surrounding the actual site of cleavage is ideally six residues long, with small side-chain amino acids in the -1 and -3 positions.

Also known in the art is the type IV signal sequence which is an example of the several types of functional signal sequences which exist in addition to the type I signal sequence detailed above. Although functionally related, the type IV signal sequence possesses a unique set of biochemical and physical attributes (Strom, M. S. and Lory, S., *J. Bacteriol.* 174, 7345-7351; 1992). These are typically six to eight amino acids with a net basic charge followed by an additional sixteen to thirty primarily hydrophobic residues. The cleavage site of a type IV signal sequence is typically after the initial six to eight amino acids at the extreme amino terminus. In addition, all type IV signal sequences contain a phenylalanine residue at the +1 site relative to the cleavage site.

Studies of the cleavage sites of twenty-six bacterial lipoprotein precursors has allowed the definition of a consensus amino acid sequence for lipoprotein cleavage. Nearly three-fourths of the bacterial lipoprotein precursors examined contained the sequence L-(A,S)-(G,A)-C at positions -3 to +1, relative to the point of cleavage (Hayashi, S. and Wu, H. C. *Lipoproteins in bacteria. J Bioenerg. Biomembr.* 22, 451-471; 1990).

It is well known that most anchored proteins found on the surface of gram-positive bacteria possess a highly conserved carboxy terminal sequence. More than fifty such proteins from organisms such as *S. pyogenes*, *S. mutans*, *E. faecalis*, *S. pneumoniae*, and others, have been identified based on their extracellular location and carboxy terminal amino acid sequence (Fischetti, V. A. Gram-positive commensal bacteria deliver antigens to elicit mucosal and systemic immunity. *ASM News* 62, 405410; 1996). The conserved region is comprised of six charged amino acids at the extreme carboxy terminus coupled to 15-20 hydrophobic amino acids presumed to function as a transmembrane domain. Immediately adjacent to the transmembrane domain is a six amino acid sequence conserved in nearly all proteins ex-

aminated. The amino acid sequence of this region is L-P-X-T-G-X, where X is any amino acid.

Amino acid sequence similarities to proteins of known function by BLAST enables the assignment of putative functions to novel amino acid sequences and allows for the selection of proteins thought to function outside the cell wall. Such proteins are well known in the art and include "lipoprotein", "periplasmic", or "antigen".

An algorithm for selecting antigenic and immunogenic *Staphylococcus aureus* polypeptides including the foregoing criteria was developed by the present inventors. Use of the algorithm by the inventors to select immunologically useful *Staphylococcus aureus* polypeptides resulted in the selection of several ORFs which are predicted to be outermembrane-associated proteins. These proteins are identified in Table 4, below, and shown in the Sequence Listing as SEQ ID NOS:5,192 to 5,255. Thus the amino acid sequence of each of several antigenic *Staphylococcus aureus* polypeptides listed in Table 4 can be determined, for example, by locating the amino acid sequence of the ORF in the Sequence Listing. Likewise the polynucleotide sequence encoding each ORF can be found by locating the corresponding polynucleotide SEQ ID in Tables 1, 2, or 3, and finding the corresponding nucleotide sequence in the sequence listing.

As will be appreciated by those of ordinary skill in the art, although a polypeptide representing an entire ORF may be the closest approximation to a protein found *in vivo*, it is not always technically practical to express a complete ORF *in vitro*. It may be very challenging to express and purify a highly hydrophobic protein by common laboratory methods. As a result, the immunologically useful polypeptides described herein as SEQ ID NOS:5,192-5,255 may have been modified slightly to simplify the production of recombinant protein, and are the preferred embodiments. In general, nucleotide sequences which encode highly hydrophobic domains, such as those found at the amino terminal signal sequence, are excluded for enhanced *in vitro* expression of the polypeptides. Furthermore, any highly hydrophobic amino acid sequences occurring at the carboxy terminus are also excluded. Such truncated polypeptides include for example the mature forms of the polypeptides expected to exist in nature.

Those of ordinary skill in the art can identify soluble portions of the polypeptide identified in Table 4, and in the case of truncated polypeptides sequences shown as SEQ ID NOS:5,192-5,255, may obtain the complete predicted amino acid sequence of each polypeptide by translating the corresponding polynucleotides sequences of the corresponding ORF listed in Tables 1, 2 and 3 and found in the sequence listing.

Accordingly, polypeptides comprising the complete amino acid of an immunologically useful polypeptide selected from the group of polypeptides encoded by the ORFs identified in Table 4, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto, and most preferably at least 99% identical thereto form an embodiment of the invention; in addition polypeptides comprising an amino acid sequence selected from the group of amino acid sequences shown in the sequence listing as SEQ ID NOS:5,191-5,255, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto and most preferably at least 99% identical thereto, form an embodiment of the invention. Polynucleotides encoding the foregoing polypeptides also form part of the present invention.

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the invention, particularly those epitope-bearing portions (antigenic regions) identified in Table 4. The epitope-bearing portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998- 4002 (1983).

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for instance, Sutcliffe, J. G., Shinnick, T. M., Green, N. and Learner, R. A. (1983) "Antibodies that react with predetermined sites on proteins", Science, 219:660-666. Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. See, for instance, Wilson et al., Cell 37:767-778 (1984) at 777.

Antigenic epitope-bearing peptides and polypeptides of the invention preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate *S. aureus* specific antibodies include: a polypeptide comprising peptides shown in Table 4 below. These polypeptide fragments have been determined to bear antigenic epitopes of indicated *S. aureus* proteins by the analysis of the Jameson-Wolf antigenic index, a representative sample of which is shown in Figure 3.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means. See, e.g., Houghten, R. A. (1985) General method for the rapid solid-phase synthesis of large numbers of peptides:

specificity of antigen-antibody interaction at the level of individual amino acids. Proc. Natl. Acad. Sci. USA 82: 5131-5135; this "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further described in U.S. Patent No. 4,631,211 to Houghten et al. (1986). Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art. See, for instance, Sutcliffe et al., *supra*; Wilson et al., *supra*;

5 Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).

10 Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. See, for instance, Geysen et al., *supra*. Further still, U.S. Patent No. 5,194,392 to Geysen (1990) describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No. 4,433,092 to Geysen (1989) describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971 to Houghten, R. A. et al. (1996) on Per-alkylated Oligopeptide Mixtures discloses linear C1-C7-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a per-alkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

15 Table 4 lists immunologically useful polypeptides identified by an algorithm which locates novel *Staphylococcus aureus* outermembrane proteins, as is described above. Also listed are epitopes or "antigenic regions" of each of the identified polypeptides. The antigenic regions, or epitopes, are delineated by two numbers x-y, where x is the number of the first amino acid in the open reading frame included within the epitope and y is the number of the last amino acid in the open reading frame included within the epitope. For example, the first epitope in ORF 168-6 is comprised of amino acids 36 to 45 of SEQ ID NO:5,192, as is described in Table 4. The inventors have identified several epitopes for each of the antigenic polypeptides identified in Table 4. Accordingly, forming part of the present invention are polypeptides comprising an amino acid sequence of one or more antigenic regions identified in Table 4. The invention further provides polynucleotides encoding such polypeptides.

20 The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are substantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity, and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

25 The invention further provides methods of obtaining homologs from other strains of *Staphylococcus aureus*, of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Staphylococcus aureus* is defined as a homolog of a fragment of the *Staphylococcus aureus* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Staphylococcus aureus* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

30 As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among those are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

35 Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS:1-5,191 or from a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ ID NOS:1-5,191 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis et al., PCR PROTOCOLS, Academic Press, San Diego, CA (1990)).

40 When using primers derived from SEQ ID NOS:1-5,191 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-5,191, one skilled in the art will recognize that by employing high stringency conditions (e.g., annealing at 50-60°C in 6X SSPC and 50% formamide, and washing at 50-65°C in 0.5X SSPC) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency

conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences which are greater than 40-50% homologous to the primer will also be amplified.

When using DNA probes derived from SEQ ID NOS:1-5,191, or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-5,191, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions (e.g., hybridizing at 50-65°C in 5X SSPC and 50% formamide, and washing at 50-65°C in 0.5X SSPC), sequences having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacteria which are closely related to *Staphylococcus aureus*.

ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and industrial purposes consistent with the type of putative identification of the polypeptide. Such identifications permit one skilled in the art to use the *Staphylococcus aureus* ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., Macmillan Publications, Ltd. NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESSES, Tramper *et al.*, Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

The various metabolic pathways present in *Staphylococcus aureus* can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS:1-5,191.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts *et al.*, *Symbiosis* 21: 79 (1986) and Voragen *et al.* in BIOCATALYSTS IN AGRICULTURAL BIOTECHNOLOGY, Whitaker *et al.*, Eds., *American Chemical Society Symposium Series* 389: 93 (1989).

The metabolism of sugars is an important aspect of the primary metabolism of *Staphylococcus aureus*. Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerase such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulonic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger *et al.*, *Biotechnology* 6(A), Rhine *et al.*, Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir *et al.*, *Biotechnology Letters* 1: 21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis *et al.*, beginning on page 357 in GENE MANIPULATIONS AND FUNGI; Benett *et al.*, Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for

quantitative determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrolysates. This application is described in Owusu *et al.*, *Biochem. et Biophysica. Acta.* 872: 83 (1986), for instance.

5 The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.*, *Biotechnology, The Textbook of Industrial Microbiology*, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose-produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, *Starch* 40:307 (1988).

10 Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See Faultman *et al.*, *Acid Proteases Structure Function and Biology*, Tang, J., ed., Plenum Press, New York (1977) and Godfrey *et al.*, *Industrial Enzymes*, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.*, *Report Industrial Enzymes by 1990*, Hel Hepner & Associates, London (1986)).

15 Another class of commercially usable proteins of the present invention are the microbial lipases, described by, for instance, Macrae *et al.*, *Philosophical Transactions of the Chiral Society of London* 310:227 (1985) and Poserke, *Journal of the American Oil Chemist Society* 61:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application 20 of lipases include the use as a detergent additive to facilitate the removal of fats from fabrics in the course of the washing procedures.

25 The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates. Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.*, *Recent Advances in the Generation of Chiral Intermediates Using Enzymes*, CRC Press, Boca Raton, Florida (1990)). The following reactions catalyzed by enzymes are of interest to organic chemists: hydrolysis of carboxylic acid esters, 30 phosphate esters, amides and nitriles, esterification reactions, trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

35 When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other hand, has been described in detail by Bud *et al.*, *Chemistry in Britain* (1987), p. 127.

40 Amino transferases, enzymes involved in the biosynthesis and metabolism of amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo-selective synthesis of only L-amino acids and generally possess uniformly high catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology* 136:479 (1987).

45 Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination. A variety of commercially important enzymes have previously been isolated from members of *Staphylococcus aureus*. These include Sau3A and Sau96I.

45 2. Generation of Antibodies

50 As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety of procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be used to generate an antibody which selectively binds the protein. Such antibodies can be either monoclonal or polyclonal antibodies, as well as fragments of these antibodies, and humanized forms.

55 The invention further provides antibodies which selectively bind to one of the proteins of the present invention and hybridomas which produce these antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., *MONOCLONAL ANTIBODY TECHNOLOGY: LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al.*, *J. Immunol. Methods* 35: 1-21 (1980), Kohler and Milstein, *Nature* 256: 495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today*

4: 72 (1983), pgs. 77-96 of Cole *et al.*, in MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc. (1985)).

5 Any animal (mouse, rabbit, *etc.*) which is known to produce antibodies can be immunized with the pseudogene polypeptide. Methods for immunization are well known in the art. Such methods include subcutaneous or interperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection.

10 The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin or galactosidase) or through the inclusion of an adjuvant during immunization.

15 For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

20 Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175: 109-124 (1988)).

25 Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

30 Techniques described for the production of single chain antibodies (U. S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

35 The present invention further provides the above-described antibodies in detectably labelled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, *etc.*), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, *etc.*) fluorescent labels (such as FITC or rhodamine, *etc.*), paramagnetic atoms, *etc.* Procedures for accomplishing such labelling are well-known in the art, for example see Sternberger *et al.*, *J. Histochem. Cytochem.* 18:315 (1970); Bayer, E. A. *et al.*, *Meth. Enzym.* 62:308 (1979); Engvall, E. *et al.*, *Immunol.* 109:129 (1972); Goding, J. W. J. *Immunol. Meth.* 13:215 (1976)).

40 The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the *Staphylococcus aureus* genome is expressed.

45 The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D. M. *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. *et al.*, *Meth. Enzym.* 34 Academic Press, N. Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immunoaffinity purification of the proteins of the present invention.

40 3. Diagnostic Assays and Kits

The present invention further provides methods to identify the expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using one of the DFs, antigens or antibodies of the present invention.

45 In detail, such methods comprise incubating a test sample with one or more of the antibodies, or one or more of the DFs, or one or more antigens of the present invention and assaying for binding of the DFs, antigens or antibodies to components within the test sample.

50 Conditions for incubating a DF, antigen or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the DF or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the DFs, antigens or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry; PCT publication WO95/32291, and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985), all of which are hereby incorporated herein by reference.

55 The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based

on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

5 In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the DFs, antigens or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound DF, antigen or antibody.

10 In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody, antigen or DF.

15 Types of detection reagents include labelled nucleic acid probes, labelled secondary antibodies, or in the alternative, if the primary antibody is labelled, the enzymatic, or antibody binding reagents which are capable of reacting with the labelled antibody. One skilled in the art will readily recognize that the disclosed DFs, antigens and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4. Screening Assay for Binding Agents

25 Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Staphylococcus aureus* fragment and contigs herein described.

In general, such methods comprise steps of:

30 (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the *Staphylococcus aureus* genome; and
 (b) determining whether the agent binds to said protein or said fragment.

35 The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

40 Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

45 In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

50 One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

55 Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, Nucl. Acids Res. 6:3073 (1979); Cooney *et al.*, Science 241:456 (1988); and Dervan *et al.*, Science 251: 1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated

to be effective in model systems. Information contained in the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

5. Pharmaceutical Compositions and Vaccines

5 The present invention further provides pharmaceutical agents which can be used to modulate the growth or pathogenicity of *Staphylococcus aureus*, or another related organism, *in vivo* or *in vitro*. As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical compositions. As used herein, the "pharmaceutical agents of the present invention" refers the pharmaceutical agents which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

10 As used herein, a pharmaceutical agent is said to "modulate the growth or pathogenicity of *Staphylococcus aureus* or a related organism, *in vivo* or *in vitro*," when the agent reduces the rate of growth, rate of division, or viability of the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth or pathogenicity by binding to an important protein thus blocking the biological activity of the protein, while other agents may bind to a component of the outer surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs 15 of the present invention and serve as a vaccine. The development and use of vaccines derived from membrane associated polypeptides are well known in the art. The inventors have identified particularly preferred immunogenic *Staphylococcus aureus* polypeptides for use as vaccines. Such immunogenic polypeptides are described above and summarized in Table 4, below.

20 As used herein, a "related organism" is a broad term which refers to any organism whose growth or pathogenicity can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a vaccine. As such, related organisms do not need to be bacterial but may be fungal or viral pathogens.

25 The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body weight daily, taking into account the routes of administration, symptoms, etc.

30 The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, etc. The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, etc. Moieties capable of mediating such effects are disclosed in, among other sources, 35 REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

40 For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal stability, biological half-life, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers 45 also may be effected in this way and can be assayed by methods well known to the skilled artisan.

45 The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (e.g., inhalation, intravenously, intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple injections.

50 In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, etc. In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from about 55 1 pg/kg to 10 mg/kg (body weight of patient), although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

55 As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each compound, or (2) the serum concentrations of each compound can

be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

5 The administration of the agent(s) of the invention may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organism's growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic administration of the compound(s) serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

10 The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

15 The agents of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, e.g., human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16th Ed., 20 Osol, A., Ed., Mack Publishing, Easton PA (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

25 Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention.

30 The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to control the duration of action by controlled release preparations is 35 to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-microcapsules and poly(methylmethacrylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES (1980).

40 The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

45 In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

6. Shot-Gun Approach to Megabase DNA Sequencing

50 The present invention further demonstrates that a large sequence can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing protocols.

55 Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present disclosure.

ILLUSTRATIVE EXAMPLESLIBRARIES AND SEQUENCING

5 1. Shotgun Sequencing Probability Analysis

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Lander and Waterman, *Genomics* 2: 231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability, P_0 , that any given base in a sequence of size L , in nucleotides, is not sequenced after a certain amount, n , in nucleotides, of random sequence has been determined can be calculated by the equation $P_0 = e^{-m}$, where $m = L/n$, the fold coverage.* For instance, for a genome of 2.8 Mb, $m=1$ when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point, $P_0 = e^{-1} = 0.37$. The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence L has not been determined and, therefore, is equivalent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage, approximately 37% of a polynucleotide of size L , in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a 2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

20 Similarly, the total gap length, G , is determined by the equation $G = Le^{-m}$, and the average gap size, g , follows the equation, $g = L/n$. Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a poly-nucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, *Genomics* 2: 231 (1988).

25 2. Random Library Construction

In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

Staphylococcus aureus DNA was prepared by phenol extraction. A mixture containing 600 ug DNA in 3.3 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 30% glycerol was sonicated for 1 min. at 0°C in a Branson 30 Model 450 Sonicator at the lowest energy setting using a 3 mm probe. The sonicated DNA was ethanol precipitated and redissolved in 500 ul TE buffer.

35 To create blunt-ends, a 100 ul aliquot of the resuspended DNA was digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200 ul BAL31 buffer. The digested DNA was phenol-extracted, ethanol-precipitated, redissolved in 100 ul TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size was excised from the gel, and the LGT agarose was melted and the resulting solution was extracted with phenol to separate the agarose from the DNA. DNA was ethanol precipitated and redissolved in 20 ul of TE buffer for ligation to vector.

40 A two-step ligation procedure was used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50 ul) contained 2 ug of DNA fragments, 2 ug pUC18 DNA (Pharmacia) cut with SmaI and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and was incubated at 14°C for 4 hr. The ligation mixture then was phenol extracted and ethanol precipitated, and the precipitated DNA was dissolved in 20 ul TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder were visualized by ethidium bromide-staining and UV illumination and identified by size as insert (i), vector (v), v+i, v+2i, v+3i, etc. The portion of the gel containing v+i DNA was excised and the v-i DNA was recovered and resuspended into 20 ul TE. The v+i DNA then was blunt-ended by T4 polymerase treatment for 5 min. at 37°C in a reaction mixture (50 ul) containing the v+i linear, 500 uM each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+i linear were dissolved in 20 ul TE. The final ligation to produce circles was carried out in a 50 ul reaction containing 5 ul of v+i linear and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture was stored at -20°C.

50 This two-stage procedure resulted in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

55 Since deviation from randomness can arise from propagation the DNA in the host, *E.coli* host cells deficient in all recombination and restriction functions (A. Greener, *Strategies* 3 (1):5 (1990)) were used to prevent rearrangements, deletions, and loss of clones by restriction. Furthermore, transformed cells were plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating was carried out as follows. A 100 ul aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) was thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7 ul aliquot of 1.42 M beta-

mercaptoethanol was added to the aliquot of cells to a final concentration of 25 mM. Cells were incubated on ice for 10 min. A 1 μ l aliquot of the final ligation was added to the cells and incubated on ice for 30 min. The cells were heat pulsed for 30 sec. at 42° C and placed back on ice for 2 min. The outgrowth period in liquid culture was eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation mixture was plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml MgCl₂ (1 M), and 1 ml MgSO₄/100 ml SOB agar. The 15 ml top layer was poured just prior to plating. Our titer was approximately 100 colonies/10 μ l aliquot of transformation.

All colonies were picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products would be deleted from the library, resulting in a slight increase in gap number over that expected.

3. Random DNA Sequencing

15

High quality double stranded DNA plasmid templates were prepared using an alkaline lysis method developed in collaboration with 5Prime → 3Prime Inc. (Boulder, CO). Plasmid preparation was performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Average template concentration was determined by running 25% of the samples on an agarose gel. DNA concentrations were not adjusted.

20

Templates were also prepared from a *Staphylococcus aureus* lambda genomic library. An unamplified library was constructed in Lambda DASH II vector (Stratagene). *Staphylococcus aureus* DNA (> 100 kb) was partially digested in a reaction mixture (200 μ l) containing 50 μ g DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23 C. The digested DNA was phenol-extracted and centrifuged over a 10- 40% sucrose gradient. Fractions containing genomic DNA of 15-25 kb were recovered by precipitation. One μ l of fragments was used with 1 μ l of DASHII vector (Stratagene) in the recommended ligation reaction. One μ l of the ligation mixture was used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract Phage were plated directly without amplification from the packaging mixture (after dilution with 500 μ l of recommended SM buffer and chloroform treatment). Yield was about 2.5x10⁹ pfu/ μ l.

25

An amplified library was prepared from the primary packaging mixture according to the manufacturer's protocol. The amplified library is stored frozen in 7% dimethylsulfoxide. The phage titer is approximately 1x10⁹ pfu/ml.

30 Mini-liquid lysates (0.1 μ l) are prepared from randomly selected plaques and template is prepared by long range PCR. Samples are PCR amplified using modified T3 and T7 primers, and Elongase Supermix (LTI).

35

Sequencing reactions are carried out on plasmid templates using a combination of two workstations (BIOMEK 1000 and Hamilton Microlab 2200) and the Perkin-Elmer 9600 thermocycler with Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers. Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. Modified T7 and T3 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are on a combination of AB 373 DNA Sequencers and ABI 377 DNA sequencers. All of the dye terminator sequencing reactions are analyzed using the 2X 9 hour module on the AB 377. Dye primer reactions are analyzed on a combination of ABI 373 and ABI 377 DNA sequencers. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

40

4. Protocol for Automated Cycle Sequencing

45

The sequencing was carried out using Hamilton Microstation 2200, Perkin Elmer 9600 thermocyclers, ABI 373 and ABI 377 Automated DNA Sequencers. The Hamilton combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-labelled sequencing primers, and reaction buffer. Reaction mixes and templates were combined in the wells of a 96-well thermocycling plate and transferred to the Perkin Elmer 9600 thermocycler. Thirty consecutive cycles of linear amplification (i.e., one primer synthesis) steps were performed including denaturation, annealing of primer and template, and extension; i.e., DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without the need for an oil overlay.

50

55 Two sequencing protocols were used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer was labelled with a different fluorescent dye, permitting the four individual reactions to be combined into one lane of the 373 or 377 DNA Sequencer for electrophoresis, detection, and base-

calling. ABI currently supplies premixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR-generated templates with both dye-primers and dye-terminators with approximately equal fidelity, although plasmid templates generally give longer usable sequences.

5 Thirty-two reactions were loaded per ABI 373 Sequencer each day and 96 samples can be loaded on an ABI 377 per day. Electrophoresis was run overnight (ABI 373) or for 2 1/2 hours (ABI 377) following the manufacturer's protocols. Following electrophoresis and fluorescence detection, the ABI 373 or ABI 377 performs automatic lane tracking and base-calling. The lane-tracking was confirmed visually. Each sequence electropherogram (or fluorescence lane trace) was inspected visually and assessed for quality. Trailing sequences of low quality were removed and the sequence 10 itself was loaded via software to a Sybase database (archived daily to 8mm tape). Leading vector polylinker sequence was removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 or ABI 377 were around 400 bp and depend mostly on the quality of the template used for the sequencing reaction.

INFORMATICS

15

1. Data Management

A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database management system and was designed to automate data flow wherever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen was based on a Unix platform, it was necessary to design and implement a variety of multi-user, client-server applications which 20 allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

2. Assembly

30 An assembly engine (TIGR Assembler) developed for the rapid and accurate assembly of thousands of sequence fragments was employed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than 10^4 fragments, the algorithm builds a hash table of 12 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. 35 Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology* 164: 765 (1988)). The contig is extended by the fragment only if strict criteria for the quality 40 of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched end, and the minimum percentage match. These criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repetitive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of alignments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information 45 coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain range of base pairs (definable for each clone based on the known clone size range for a given library).

3. Identifying Genes

50

The predicted coding regions of the *Staphylococcus aureus* genome were initially defined with the program zorf, which finds ORFs of a minimum length. The predicted coding region sequences were used in searches against a database of all *Staphylococcus aureus* nucleotide sequences from GenBank (release 92.0), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence 55 matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and compared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs of at least 80 amino acids that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases.

ORFs of at least 120 amino acids that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3.

ILLUSTRATIVE APPLICATIONS

1. Production of an Antibody to a *Staphylococcus aureus* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can be chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows.

2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., *Nature* 256:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., *Meth. Enzymol.* 70:419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.* *Basic Methods in Molecular Biology* Elsevier, New York. Section 21-2 (1989).

3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in: *Handbook of Experimental Immunology*, Wier, D., ed., Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980).

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, they are useful in various animal models of Staphylococcal disease known to those of skill in the art as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a potential immunotherapeutic reagent.

3. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Staphylococcus aureus* genome, such as those of Tables 1-3 and SEQ ID NOS:1-5,191 can be used, in accordance with the present invention, to prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approxi-

mately the same. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

4. Gene expression from DNA Sequences Corresponding to ORFs

5 A fragment of the *Staphylococcus aureus* genome provided in Tables 1-3 is introduced into an expression vector using conventional technology. Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U. S. Patent No. 5,082,767, incorporated herein by this reference.

10 The following is provided as one exemplary method to generate polypeptide(s) from cloned ORFs of the *Staphylococcus aureus* genome fragment. Bacterial ORFs generally lack a poly A addition signal. The addition signal sequence can be added to the construct by, for example, splicing out the poly A addition sequence from pSG5 (Stratagene) using BglI and Sall restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene) for use in eukaryotic expression systems. pXT1 contains the LTRs and a portion of the gag gene of Moloney Murine Leukemia Virus. The positions of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The *Staphylococcus aureus* DNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the *Staphylococcus aureus* DNA and containing restriction endonuclease sequences for PstI incorporated into the 5' primer and BglII at the 5' end of the corresponding *Staphylococcus aureus* DNA 3' primer, taking care to ensure that the *Staphylococcus aureus* DNA is positioned such that its followed with the poly A addition sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with BglII, purified and ligated to pXT1, now containing a poly A addition sequence and digested BglII.

15 20 25 The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface. Since it may be necessary to purify and locate the transfected product, 30 35 synthetic 15-mer peptides synthesized from the predicted *Staphylococcus aureus* DNA sequence are injected into mice to generate antibody to the polypeptide encoded by the *Staphylococcus aureus* DNA.

40 Alternatively and if antibody production is not possible, the *Staphylococcus aureus* DNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as, for example, a globin fusion. Antibody to the globin moiety then is used to purify the chimeric protein. Corresponding protease cleavage sites are engineered between the globin moiety and the polypeptide encoded by the *Staphylococcus aureus* DNA so that the latter may be freed from the formed by simple protease digestion. One useful expression vector for generating globin chimerics is pSG5 (Stratagene). This vector encodes a rabbit globin. Intron II of the rabbit globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, cited elsewhere herein, and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptides of the invention also may be produced using *in vitro* translation systems such as *in vitro* ExpressTM Translation Kit (Stratagene).

45 While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

All patents, patent applications and publications referred to above are hereby incorporated by reference.

S. aureus - Coding regions containing known sequences

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	Percent HSP nt ident	ORF nt length
1	1	1419	757	[emb]X17301 S41D	[S. aureus] DNA for hid gene and for part of agr gene	100	663
1	2	3373	2452	[emb]X52243 SAG	[S. aureus] agrA, agrB and hid genes	99	409
1	5	6418	5651	[db]D14711 STAH	[Staphylococcus aureus] HSP10 and HSP60 genes	98	223
5	1	807	439	[emb]X72700 IS41V	[S. aureus] genes for S and F components of Panton-Valentine leucocidins	81	216
5	4	5031	3571	[emb]X72700 IS41V	[S. aureus] genes for S and F components of Panton-Valentine leucocidins	95	424
10	1	86	904	[gb]L25288	[Staphylococcus aureus] gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	98	715
16	5	5302	6246	[gb]U35773	[Staphylococcus aureus] prolipoprotein diacylglycerol transferase (tgt) gene, complete cds	94	251
16	6	6249	7091	[gb]U35773	[Staphylococcus aureus] prolipoprotein diacylglycerol transferase (tgt) gene, complete cds	99	843
16	7	7084	7584	[gb]U35773	[Staphylococcus aureus] prolipoprotein diacylglycerol transferase (tgt) gene, complete cds	99	342
20	1	995	549	[gb]U19300	[Staphylococcus aureus] DNA sequence encoding three ORFs, complete cds; prophase phi-11 sequence homology, 5' flank	100	443
20	2	1011	841	[gb]U19300	[Staphylococcus aureus] DNA sequence encoding three ORFs, complete cds; prophase phi-11 sequence homology, 5' flank	91	137
20	3	2010	1798	[gb]U19300	[Staphylococcus aureus] DNA sequence encoding three ORFs, complete cds; prophase phi-11 sequence homology, 5' flank	100	110
20	4	5300	3825	[gb]M6714	[Staphylococcus aureus] peptide-glycan hydrolase gene, complete cds	100	948
20	5	4788	4232	[gb]M714	[Staphylococcus aureus] peptide-glycan hydrolase gene, complete cds	100	309
26	1	2	145	[gb]U10721	[Staphylococcus aureus] isoleucyl-tRNA synthetase (iles) gene, partial cds	100	126
26	1	2	84	[gb]U10721	[Staphylococcus aureus] isoleucyl-tRNA synthetase (iles) gene, partial cds	99	430
26	3	763	3531	[emb]X74219 S41V	[S. aureus] gene for isoleucyl-tRNA synthetase	99	2769
29	3	1261	4392	[gb]U66651	[Staphylococcus aureus] DNA fragment with class II promoter activity	100	117
31	14	14977	133463	[emb]X173089 S41P	[S. aureus] genes P1 and P2	99	1351
31	15	14241	13855	[emb]X73089 S41P	[S. aureus] genes P1 and P2	98	258
38	17	14284	133112	[gb]M127151	[S. aureus] gdh gene encoding lipase (glycerol ester hydrolase)	100	372
38	19	133434	135518	[gb]P27151	[S. aureus] gdh gene encoding lipase (glycerol ester hydrolase)	100	2085

TABLE 1

S. aureus - Coding regions containing known sequences

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
46	2	514	1727	gb U7374	Staphylococcus aureus type 8 capsule genes, capA, capB, capC, capD, capE, capF, capG, capH, capI, capJ, capK, capL, capM, capN, capO, capP, complete cds	98	1209	1209
46	3	1720	2295	gb U7374	Staphylococcus aureus type 8 capsule genes, capA, capB, capC, capD, capE, capF, capG, capH, capI, capJ, capK, capL, capM, capN, capO, capP, complete cds	98	576	576
46	4	2259	3182	gb U7374	Staphylococcus aureus type 8 capsule genes, capA, capB, capC, capD, capE, capF, capG, capH, capI, capJ, capK, capL, capM, capN, capO, capP, complete cds	97	924	924
46	5	3173	4498	gb U7374	Staphylococcus aureus type 8 capsule genes, capA, capB, capC, capD, capE, capF, capG, capH, capI, capJ, capK, capL, capM, capN, capO, capP, complete cds	98	1283	1226
46	6	4536	5720	gb U7374	Staphylococcus aureus type 8 capsule genes, capA, capB, capC, capD, capE, capF, capG, capH, capI, capJ, capK, capL, capM, capN, capO, capP, complete cds	98	1185	1185
46	7	6455	6120	gb U7374	Staphylococcus aureus type 8 capsule genes, capA, capB, capC, capD, capE, capF, capG, capH, capI, capJ, capK, capL, capM, capN, capO, capP, complete cds	99	278	236
48	1	2	955	gb J23693	Staphylococcus aureus rckA gene, complete cds	99	954	954
50	3	4465	2824	emb X55039 SAFN	S.aureus AhpC gene	100	88	1542
50	4	4108	3515	emb X55039 SAFN	S.aureus AhpC gene	98	540	594
54	3	5074	3592	emb K2292 SAFN	S.aureus fnbB gene for fibronectin binding protein B	100	1668	1663
54	4	4865	4122	emb X2292 SAFN	S.aureus fnbB gene for fibronectin binding protein B	99	720	744
54	5	5056	4562	emb K62992 SAFN	S.aureus fnbB gene for fibronectin binding protein B	100	463	495
54	6	13386	8300	gb J0151	S.aureus fibronectin-binding protein (fnbA) gene, complete cds	100	3087	3087
58	1	1743	2619	emb XH7104 SADN	S.aureus mdr, pbp4 and tagD genes (SG511-55 isolate)	89	68	107
58	4	2858	3280	emb X97106 SAPB	S.aureus abcA, pbp4, and tagD genes	99	423	423
58	5	6005	4701	emb X97106 SAPB	S.aureus abcA, pbp4, and tagD genes	99	1305	1305
58	6	5677	5378	gb J29478	Staphylococcus aureus ABC transporter-like protein abcA (abcA) gene, partial cds	100	300	300
58	7	5086	6640	emb X97106 SAPB	S.aureus abcA, pbp4, and tagD genes	99	1755	1755
72	1	888	445	gb J21854	S.aureus agr gene encoding an accessory gene regulator protein, complete cds	100	444	444
72	2	2457	1453	emb X52543 SAAG	S.aureus agrA, agrB and hid genes	99	673	1005

TABLE 1

S. aureus - Coding regions containing known sequences

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	HSP nc	percent	ORF nc length	HSP nc length
82	1	357	3917	emb X64172 SARP	<i>S. aureus</i> rpl1, rpl20, rps11f and rps12 genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta L beta' chains	99	2336	3561	5
82	2	4027	7677	emb X69233 SARP	<i>S. aureus</i> DNA for rps1 gene	99	3171	3651	10
82	3	7745	8068	gb U20869	<i>Staphylococcus aureus</i> ribosomal protein S7 (rpsC) and ORF 11 genes, partial cds	100	320	324	15
82	4	8103	8579	gb U20869	<i>Staphylococcus aureus</i> ribosomal protein S7 (rpsC) and ORF 1 genes, partial cds	100	477	477	20
82	5	8618	8821	gb U20869	<i>Staphylococcus aureus</i> ribosomal protein S7 (rpsC) gene, complete cds, ribosomal protein S7 (rpsC) and ORF 1 genes, partial cds	100	154	204	25
84	1	18	191	gb U73374	<i>Staphylococcus aureus</i> type 8 capsule genes, cap8A, cap8C, cap8D, cap8E, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	164	174	30
84	2	189	893	gb U73374	<i>Staphylococcus aureus</i> type 8 capsule genes, cap8A, cap8C, cap8D, cap8E, cap8G, cap8H, cap8I, cap8K, cap8M, cap8N, cap8O, cap8P, complete cds	94	705	705	35
84	3	887	1660	gb U73374	<i>Staphylococcus aureus</i> type 8 capsule genes, cap8A, cap8C, cap8D, cap8E, cap8G, cap8H, cap8I, cap8K, cap8M, cap8N, cap8O, cap8P, complete cds	99	774	774	40
84	4	1584	3503	gb U73374	<i>Staphylococcus aureus</i> type 8 capsule genes, cap8A, cap8C, cap8D, cap8E, cap8G, cap8H, cap8I, cap8K, cap8M, cap8N, cap8O, cap8P, complete cds	98	1920	1920	45
84	5	3394	4521	gb U73374	<i>Staphylococcus aureus</i> type 8 capsule genes, cap8A, cap8C, cap8D, cap8E, cap8G, cap8H, cap8I, cap8K, cap8M, cap8N, cap8O, cap8P, complete cds	97	1128	1128	50
84	6	4519	5643	gb U73374	<i>Staphylococcus aureus</i> type 8 capsule genes, cap8A, cap8C, cap8D, cap8E, cap8G, cap8H, cap8I, cap8K, cap8M, cap8N, cap8O, cap8P, complete cds	97	1125	1125	55
96	1	1245	3896	emb 218852 SACF	<i>S. aureus</i> gene for clumping factor	83	660	2652	60
97	2	625	882	gb U41072	<i>[Staphylococcus aureus</i> tRNA synthetase (tRNA genes, partial cds, complete cds)]	97	68	258	65
111	1	3	452	gb L41499	<i>[Staphylococcus aureus</i> ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds]	100	450	450	70
111	2	526	1041	gb L41499	<i>[Staphylococcus aureus</i> ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds]	99	516	516	75
117	2	1278	1958	gb H83994	<i>[Staphylococcus aureus</i> proline/protein signal peptidase (lsp) gene, complete cds]	100	61	681	80

TABLE 1

S. aureus - Coding regions containing known sequences

Contig	ORF ID	Start (nt)	Stop (nt)	match erosion	match gene name	percent HSP	HSP nt	ORF nt
	ID	(nt)	(nt)			Ident	length	length
118	4	3787	4254	[gb U30690 STAN	[Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP10; ORF35; complete	99	467	468
130	4	2597	3640	[emb X13290 SATN	[Staphylococcus aureus multi-resistance plasmid pSK1 DNA containing	78	936	1044
130	5	3813	4265	[emb Z16422 SAD1	[S. aureus dfr8 gene for dihydrofolate reductase	98	416	453
130	6	4109	5172	[emb Z16422 SAD1	[S. aureus dfr8 gene for dihydrofolate reductase	98	607	864
136	4	5296	6207	[emb X71437 SACY	[S. aureus genes gyrB, gyrA and recF (partial)	97	838	912
136	5	11680	8987	[gb D10489 STAG	[Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	2694	2694
136	6	12886	10900	[gb D10489 STAG	[Staphylococcus aureus genes for DNA gyrase A and B, complete cds	99	1947	1947
136	7	12392	11175	[gb S77055	[recF cluster: dna-replisome assembly protein...gyrB/DNA gyrase beta subunit [Staphylococcus aureus, YB816, Genomic, 5 genes, 1573 nt]	99	822	828
143	3	4171	2867	[gb U36779	[Staphylococcus aureus S-adenosylmethionine synthetase gene, complete cds	99	1105	1105
143	4	3100	4281	[gb LA2943	[Staphylococcus aureus ftcne KIN50] phosphoenolpyruvate carboxykinase (pcckA) gene, complete cds	100	1170	1182
143	5	4254	4718	[gb U51131	[Staphylococcus aureus phosphoenolpyruvate carboxykinase (pcckA) gene, complete cds	100	449	465
143	9	6377	7281	[gb U51132	[Staphylococcus aureus o-succinylbenzolic acid CoA ligase (menE) and o-succinylbenzoic acid synthetase (menC) genes, complete cds	100	75	285
143	10	9464	8361	[gb U51132	[Staphylococcus aureus o-succinylbenzoic acid CoA ligase (menE) and o-succinylbenzoic acid synthetase (menC) genes, complete cds	100	1104	1104
143	11	11232	9768	[gb U51132	[Staphylococcus aureus o-succinylbenzoic acid CoA ligase (menE) and o-succinylbenzoic acid synthetase (menC) genes, complete cds	100	1485	1485
143	12	10739	10310	[gb U51132	[Staphylococcus aureus o-succinylbenzoic acid CoA ligase (menE) and o-succinylbenzoic acid synthetase (menC) genes, complete cds	100	332	420
152	5	2454	3437	[emb X58314 SARD	[S. aureus pdhB, pdhC and pdhD genes for soyate decarboxylase, dihydrolipopamide acetyltransferase and dihydrolipopamide dehydrogenase	99	305	984
152	6	1511	4820	[emb X58314 SARD	[S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipopamide acetyltransferase and dihydrolipopamide dehydrogenase	98	1308	1308
152	7	4818	6230	[emb X58314 SARD	[S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipopamide acetyltransferase and dihydrolipopamide dehydrogenase	99	1413	1413
153	1	387	1536	[gb S77055	[recF cluster: dna-replisome assembly protein...gyrB/DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	100	276	276
153	2	1677	2152	[gb S77055	[recF cluster: dna-replisome assembly protein...gyrB/DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	99	1140	1140

TABLE 1

S. aureus - Coding regions containing known sequences

Contig	ORF	Start	Stop	match	seqch gene name	Percent ISP nt Ident	ORF nt length
10	ORF10	(Int)	(Int)	accession			
153	3	2143	2269	gb S77055	recf cluster: diaA+replisome assembly protein...gyrB-DNA gyrase beta subunit (Staphylococcus aureus, YBB16, Genomic, 5 genes, 3573 nt)	99	113
154	10	10792	9314	gb U06451	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	91	154
154	11	9935	9615	gb U06451	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	99	229
154	12	9943	10167	gb U06451	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	94	123
154	13	10089	11501	gb U06451	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	99	1326
159	1	2195	1212	gb D028879 STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	71
161	3	2596	2270	gb M433934	Staphylococcus aureus prolineprotein signal peptidase (lsp) gene, complete cds	92	203
162	1	1406	705	gb U021221	Staphylococcus aureus hyaluronate lyase (hyalA) gene, complete cds	100	702
163	4	1263	1772	gb U9770	Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds	96	127
164	7	4774	9117	gb J086727 D867	Staphylococcus aureus DNA for DNA polymerase III, complete cds	99	3470
168	7	7448	6447	gb U21636	Staphylococcus aureus cmp-binding-factor 1 (cbf1) and ORF X genes, complete cds	100	1002
168	8	9538	7961	gb U21636	Staphylococcus aureus cmp-binding-factor 1 (cbf1) and ORF X genes, complete cds	99	1158
173	6	9240	7801	gb J03479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacZ) genes, complete cds	100	1440
173	7	111232	9322	gb U03479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacZ) genes, complete cds	99	1731
173	8	8235	6704	gb J03479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacZ) genes, complete cds	100	420
173	9	10168	9839	gb J03479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacZ) genes, complete cds	100	310
173	10	111815	110829	gb X14627 SALA	Staphylococcus aureus lacC and lacD genes	100	987
173	11	112721	111774	gb X14627 SALA	Staphylococcus aureus lacC and lacD genes	100	948
173	12	112838	112305	gb M64724	S. aureus tagatose 6-phosphate isomerase gene, complete cds	100	534
173	13	113233	112773	gb U132103	Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA1) partial cds	100	471
173	14	114633	113866	gb U132103	Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA1) partial cds	100	768

TABLE 1

5
10
15
20
25
30
35
40
45
50

S. aureus - Coding regions containing known sequences

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent HSP nt ident	ORF nt length
178	1	2	655	[gb U52961]	[Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds]	100	115
178	2	2201	1482	[gb U52961]	[Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds]	100	720
178	3	2361	1909	[gb U52961]	[Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds]	100	453
178	4	1551	1853	[gb U52961]	[Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds]	100	303
178	5	3541	2777	[gb U29451]	[Staphylococcus aureus lys and lyc genes, complete cds]	99	765
178	6	3294	3025	[gb U29451]	[Staphylococcus aureus lys and lyc genes, complete cds]	99	270
181	1	1114	590	[gb H01177]	[S. aureus sigma factor (lpha) gene, complete cds]	99	499
182	1	3	341	[emb X61307 SAST]	[Staphylococcus aureus spa gene for protein A]	98	217
182	2	690	2112	[gb U01786]	[S. aureus spa gene coding for protein A, complete cds]	97	1322
182	3	5861	4251	[emb X61307 SASP]	[Staphylococcus aureus spa gene for protein A]	99	119
185	1	3	824	[gb U11979]	[Staphylococcus aureus chorismate synthase (arcC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydrodauvinate synthase (arcB) and geranylgeranyl pyrophosphate synthase homolog (garC) genes, partial cds]	90	132
191	3	841	2760	[emb X17679 SACO]	[Staphylococcus aureus coag gene for coagulase]	99	1920
191	4	2967	3143	[emb X116157 SAST]	[Staphylococcus aureus gene for staphylococcal kinase (indk) gene, complete cds, dehydrodauvinate synthase (arcB) and geranylgeranyl pyrophosphate synthase homolog (garC) genes, partial cds]	99	177
191	5	5168	4566	[emb X116157 SAST]	[Staphylococcus aureus gene for staphylococcal kinase (indk) gene, complete cds, dehydrodauvinate synthase (arcB) and geranylgeranyl pyrophosphate synthase homolog (garC) genes, partial cds]	99	250
196	1	1741	872	[gb U116472]	[Staphylococcus aureus lys-tRNA synthetase gene, complete cds, transfer RNA (tRNA) genes, 5S ribosomal RNA (5S rRNA) gene, 15S ribosomal RNA (15S rRNA) gene, 23S ribosomal RNA (23S rRNA) gene]	99	870
198	3	1688	2011	[emb X91305 SAPP]	[S. aureus pspH and pspI genes]	99	324
198	4	2005	2110	[emb X91305 SAPP]	[S. aureus pspH and pspI genes]	97	304
202	1	163	1305	[emb X917985 SA12]	[S. aureus otts 1,2,3 & 4]	99	1143
202	2	1303	2175	[emb X73089 SA1]	[S. aureus genes P1 and P2]	94	444
210	1	3114	1558	[gb U017766 STNA]	[Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs]	99	1552
210	2	2939	2232	[gb U41499]	[Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds]	99	708
214	11	7439	7770	[db D86240 D862]	[Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds]	96	157
214	12	7440	7771	[db D86240 D863]	[Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds]	96	142

TABLE 1

S. aureus - Coding regions containing known sequences

contig	orf ID	start	stop	match accession	match gene name	percent HSP int	HSP int	ORF nt length
	ID	int	int			ident		
216	1	398	1118	embX22700 SAPV	S. aureus genes for S and F components of Panton-Valentine leucocidins	88	265	921
219	2	1810	1073	dbj D0690 STAN	[Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35; complete	100	60	738
219	3	2979	2035	dbj D0690 STAN	[Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35; complete	99	945	945
219	4	4359	3196	dbj D0690 STAN	[Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35; complete	99	1164	1164
219	5	7044	116	dbj D0690 STAN	[Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35; complete	98	1869	1869
219	6	6557	5883	dbj D0690 STAN	[Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35; complete	99	675	675
219	7	6801	6334	dbj D0690 STAN	[Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35; complete	98	468	468
221	8	10816	10034	gb 119398	[Staphylococcus aureus phosphatidylinositol-specific phospholipase C (pic)	91	67	783
223	1	2855	1566	gb U73374	[Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8D, cap8E, cap8G, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8H, cap8I, cap8P, cap8R, complete cds	99	102	1150
234	1	2	1357	embX97985 SA12	[S. aureus orf8 1,2,3 & 4	100	176	1356
234	2	1694	2485	embX97985 SA12	[S. aureus orf8 1,2,3 & 4	100	792	792
234	3	2648	3146	embX97985 SA12	[S. aureus orf8 1,2,3 & 4	99	501	501
234	4	3120	4604	embX97985 SA12	[S. aureus orf8 1,2,3 & 4	99	1305	1485
236	6	3826	5332	gb U48826	[Staphylococcus aureus elastin binding protein (abp5) gene, complete cds	96	648	1497
248	1	2	403	embX62288 SPE	[S. aureus DNA for penicillin-binding protein 2	100	103	402
248	2	388	852	gb J35426	[Staphylococcus aureus penicillin-binding protein 2 (ppb2) gene, complete	99	465	465
253	2	1519	1093	gb U46541	[Staphylococcus aureus sara gene, complete cds	96	447	447
254	1	150	1835	gb U57601	[Staphylococcus aureus scca gene, complete cds	94	142	1686
254	3	1973	2728	gb U57601	[Staphylococcus aureus sccA gene, complete cds	99	756	756
260	1	2	1900	gb U48069	[Staphylococcus aureus glycosidase II (glp) gene, complete cds	99	1213	1699
265	1	1	942	dbj D21111 STAS	[Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	99	941	942

TABLE 1

S. aureus - Coding regions containing known sequences

Contig	ORF ID	Start int	Stop int	match accession	match gene name	Percent LSP nt ident	ORF nt length
265	2	688	476	[dbj]021131 STAS	[Staphylococcus aureus gene for a participant in homogenous expression of high-level methicillin resistance, complete cds	99	213
265	3	2418	1765	[dbj]021131 STAS	[Staphylococcus aureus gene for a participant in homogenous expression of high-level methicillin resistance, complete cds	98	69
266	1	2	1018	[dbj]014711 STAH	[Staphylococcus aureus HSP10 and HSP60 genes	98	743
282	1	1	525	[gbf]72488 nt	[heat-shock/bilinogen synthase (Staphylococcus aureus, SA1939, Genomic, 1087	100	110
282	2	516	1502	[gbf]572488 nt	[heat-shock/bilinogen synthase (Staphylococcus aureus, SA1939, Genomic, 1087	100	952
284	1	3	170	[gbf]963176 nt	[Staphylococcus aureus helicase required for τ 181 replication (pcrA) gene, complete cds	98	84
284	2	282	1034	[gbf]963176 nt	[Staphylococcus aureus helicase required for τ 181 replication (pcrA) gene, complete cds	100	712
284	3	1028	2026	[gbf]963176 nt	[Staphylococcus aureus helicase required for τ 181 replication (pcrA) gene, complete cds	99	999
284	4	1990	2202	[gbf]963176 nt	[Staphylococcus aureus helicase required for τ 181 replication (pcrA) gene, complete cds	98	187
289	1	1516	1951	[gbf]912470 nt	[S. aureus Sa1Al-restriction-enzyme and Sa1Al-modification-enzyme genes, complete cds	99	338
303	1	2	868	[gbf]01055 nt	[Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds	99	867
303	2	1409	2363	[gbf]01055 nt	[Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds	100	975
303	3	2367	3161	[gbf]01055 nt	[Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds	99	793
305	1	2707	1355	[dbj]D12366 STAA	[Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	1343
311	1	2628	1315	[gbf]42945 nt	[Staphylococcus aureus lys and lyt genes, complete cds	98	1314
312	6	7019	7870	[gbf]114017 nt	[Staphylococcus aureus methicillin-resistance protein (macR) gene and unknown ORF, complete cds	74	351
323	1	1998	1003	[gbf]031175 nt	[Staphylococcus aureus D-specific D-2-hydroxyacid dehydrogenase (ddh) gene, complete cds	98	996
326	1	1	237	[emb]Y0035 SASP	[Staphylococcus aureus V8 serine protease gene	100	237
338	1	687	388	[emb]X63389 SALE	[S. aureus leuF-P83 gene for F component of leucocidin R	98	259
338	2	1828	1088	[emb]Y63389 SALE	[S. aureus leuF-P81 gene for F component of leucocidin R	1	97
338	2	1828	1088	[emb]Y63389 SALE	[S. aureus leuF-P81 gene for F component of leucocidin R	1	137
338	2	1828	1088	[emb]Y63389 SALE	[S. aureus leuF-P81 gene for F component of leucocidin R	1	741

TABLE 1

S. aureus - Coding regions containing known sequences

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ISP nt ident	ORF nt length
342	1	2	579	1754 [gb U06462]	[Staphylococcus aureus SAM FtsZ (ftsZ) gene, complete cds]	100	1176
344	2	2	517	1248 [emb U01281 S4NU]	[S. aureus mRNA for nucleic acid	98	732
349	1	1	457	230 [gb HM20393]	[S. aureus bacteriophage phi-11 attachment site (attB)]	96	172
353	1	1	1016	516 [gb HM3994]	[Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds]	100	187
353	2	1582	1046	[gb HM3994]	[Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds]	99	537
356	1	1	3	674 [gb U0503]	[Staphylococcus aureus MHC class II analog gene, complete cds]	75	672
361	1	1	903	[gb U19298]	[Staphylococcus aureus phosphatidylinositol-specific phospholipase C (plc) gene, complete cds]	98	747
361	2	1103	1507 [gb U19298]	[Staphylococcus aureus phosphatidylinositol-specific phospholipase C (plc) gene, complete cds]	97	68	
373	1	1	3	1148 [emb X62288 SAFE]	[S. aureus DNA for penicillin-binding protein 2]	99	1146
389	1	3	1904	1248 [emb X62282 SAYS]	[S. aureus target site DNA for IS431 insertion	97	349
400	1	1	1	540 [emb X61716 S4NU]	[S. aureus hly gene encoding sphingomyelinase]	99	389
400	2	1	1693	1187 [emb XJ34041 S4NU]	[Staphylococcus aureus atl gene for beta-hemolysin]	99	178
408	1	1810	1049 [gb S76213]	[lsp23-alkaline shock protein 23 (methicillin-resistant) [Staphylococcus aureus, 912, Genomic, 1360 nt]	99	163	
416	1	1	2	217 [gb U41499]	[Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds]	100	216
416	2	1	854	639 [dbj D17366 STNA]	[Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs]	100	186
421	2	1	1262	2509 [gb U43099]	[Transposon Tn5404 and insertion sequences IS1111 and IS1112 (from Staphylococcus aureus) DNA]	99	1240
422	1	1	2	325 [gb K02985]	[S. aureus (strain RN420) transposon Tn534 insertion site]	96	200
427	1	1	865	434 [dbj D28879 STAP]	[Staphylococcus aureus gene for penicillin-binding protein 1, complete cds]	100	432
427	2	1	1829	1122 [dbj D28819 STAP]	[Staphylococcus aureus gene for penicillin-binding protein 1, complete cds]	100	151
435	1	2	808	[dbj D86240 D862]	[Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds]	100	556
435	2	2	832	999 [dbj D86240 D862]	[Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds]	100	134
436	1	1	1341	685 [emb X1768 SAFE]	[S. aureus factor essential for expression of methicillin resistance (fma) gene, complete cds, and trpA gene, 3' end]	97	657

TABLE 1

5
10
15
20
25
30
35
40
45
50
55

S. aureus - Coding regions containing known sequences

Contig	ORF ID	Start (nt)	Stop (nt)	match	gene name	percent HSP nt	ORF nt
	ID					Ident	length
436	2	2403	1657	emb X176B SAFE	S. aureus factor essential for expression of methicillin resistance (femA)	100	294
					S. aureus genes for S and F components of Panton-Valentine leucocidins	84	204
442	1	347	1100	emb X272001 SAV	[Staphylococcus aureus gamma-hemolysin components A, B and C (hla, hlb, hlc) genes, complete cds]	98	187
445	2	1906	2178	gb U01055	[Staphylococcus aureus gamma-hemolysin components A, B and C (hla, hlb, hlc) genes, complete cds]	100	51*
447	1	167	1078	gb U19770	[Staphylococcus aureus pyrocidone carboxyl peptidase (pcp) gene, complete cds]	96	597
					[Staphylococcus aureus pyrocidone carboxyl peptidase (pcp) gene, complete cds]	75	609
447	2	1176	1784	gb U19770	[Staphylococcus aureus gamma-hemolysin components A, B and C (hla, hlb, hlc) genes, complete cds]	99	2418
454	3	7309	4319	emb 218952 SAFE	S. aureus gene for clumping factor	99	633
472	4	7896	5479	gb U25288	[Staphylococcus aureus gyrase-like protein alpha and beta subunit (gyrA and gyrB) genes, complete cds]	100	2418
472	5	8120	6792	gb U25288	[Staphylococcus aureus gyrase-like protein alpha and beta subunit (gyrA and gyrB) genes, complete cds]	99	1328
475	2	566	889	emb X2543 SAAC	[S. aureus agrA, agrB and hld genes]	100	76
481	4	1922	1560	emb X64172 SARP	[S. aureus rplB, rpsB2, rpsB3, rpsC genes for ribosomal protein L7/L12, hypothetical protein ORF20, DNA-directed RNA polymerase beta 4 beta chains]	100	250
481	5	1244	1534	emb X64172 SARP	[S. aureus rplB, rpsB2, rpsB3, rpsC genes for ribosomal protein L7/L12, hypothetical protein ORF20, DNA-directed RNA polymerase beta 4 beta chains]	100	224
487	2	1388	1188	gb UH3994	[Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds]	98	72
489	1	2377	1370	gb U21221	[Staphylococcus aureus hyaluronate lyase (hyal) gene, complete cds]	99	1368
503	2	1135	653	gb UH3994	[Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds]	100	108
511	3	1613	2242	gb U14017	[Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds]	84	423
511	4	3122	2700	gb S76213	[Staphylococcus aureus shock protein 23 (methicillin resistance) [Staphylococcus aureus, 91, genomic, 1360 nt]	99	323
520	2	1758	1297	emb X7204 SAF1	[S. aureus fib gene for fibrinogen-binding protein]	99	540
520	3	1436	1801	emb X72013 SAF1	[S. aureus fib gene for fibrinogen-binding protein]	99	221
526	1	2150	1092	lib U17366 STAA	[Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs]	99	421
						75	1059

TABLE 1

5
10
15
20
25
30
35
40
45
50
55

S. aureus - Coding regions containing known sequences

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent IISP nt ident	IISP nt length	ORF nt length
528	2	58	963	[gb]U93001	[Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11' sequence homology, 5' flank]	99	260	906
528	3	1098	2870	[gb]U93001	[Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11' sequence homology, 5' flank]	99	866	1773
530	1	3	434	[gb]U319791	[Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroaunate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (garCC) genes, partial cds]	99	432	432
530	2	1211	2395	[gb]U319791	[Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroaunate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (garCC) genes, partial cds]	91	1185	1185
530	3	2409	2801	[gb]U319791	[Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroaunate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (garCC) genes, partial cds]	88	181	193
530	4	2690	3484	[gb]U050041	[Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' and cds; 3'-phosphohikimate-1-carboxyvinyltransferase (aroC) gene, complete cds; ORF, complete cds]	100	75	795
530	5	3482	4792	[gb]U050041	[Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' and cds; 3'-phosphohikimate-1-carboxyvinyltransferase (aroC) gene, complete cds; ORF, complete cds]	99	905	1311
530	6	4790	5380	[gb]U050041	[Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' and cds; 3'-phosphohikimate-1-carboxyvinyltransferase (aroC) gene, complete cds; ORF, complete cds]	100	196	591
539	1	3	338	[emb]X76490 SAGL S.aureus (bb270) glnA and glnR genes		99	336	336
539	2	336	527	[emb]X76490 SAGL S.aureus (bb270) glnA and glnR genes		100	189	192
534	1	727	365	[gb]U333741	[Staphylococcus aureus type 8 capsule genes, capA8, capB8, capC8, capD8, capE, capF, capG, capH, capI, capJ, capK, capL, capM, capN, capO, capP, complete cds]	100	54	363
534	2	2175	1252	[gb]U333741	[Staphylococcus aureus type 8 capsule genes, capA, capB, capC, capD, capE, capF, capG, capH, capI, capJ, capK, capL, capM, capN, capO, capP, complete cds]	99	916	924
534	3	1574	1374	[gb]U333741	[Staphylococcus aureus type 8 capsule genes, capA, capB, capC, capD, capE, capF, capG, capH, capI, capJ, capK, capL, capM, capN, capO, capP, complete cds]	96	122	201
584	2	1019	705	[gb]U312211	[Staphylococcus aureus hyaluronate lyase (hylA) gene, complete cds]	99	306	315
587	3	1475	4288	[emb]Z18852 SAGF S.aureus gene for penicillin-binding protein 1, complete cds		98	2588	2814
588	1	2881	1953	[dub]028379 STAP [Staphylococcus aureus DNA sequence for clumping factor		99	1873	1929

TABLE 1

S. aureus - Coding regions containing known sequences

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	ISP nc length	ORF nc length	
605	1	2	745	gb DR6240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltU genes, complete cds	98	138	744	
609	1	1626	816	emb K27690 ISAU	S. aureus (gb270) glnA and glnR genes	100	495	813	
614	1	1280	642	gb M2103	Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds	99	639	639	
626	1	2508	1255	gb M83176	Staphylococcus aureus helicase required for T181 replication (pctA) gene, complete cds	100	225,	1254	
626	2	3115	2284	gb M63176	Staphylococcus aureus helicase required for T181 replication (pctA) gene, complete cds	99	938	1032	
629	1	1959	1001	emb X17688 SAFE	S. aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	99	990	999	
629	2	1407	1195	emb X17688 SAFE	S. aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	98	194	213	
631	1	2	5126	2228	emb 218852 SAFC	S. aureus gene for clumping factor	82	489	1899
632	1	3	551	emb 230568 ISAT	S. aureus (RN4220) genes for potential ABC transporter and potential membrane spanning protein	99	549	549	
632	2	529	132	emb 230568 ISAT	S. aureus (RN4220) genes for potential ABC transporter and potential membrane spanning protein	99	795	795	
651	1	1909	1070	gb U19300	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds, prophage phi-1 sequence homology, 5' flank	99	478	840	
657	2	1800	1105	gb U14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	84	456	696	
662	1	908	456	emb X13404 SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	100	369	453	
663	2	230	475	emb X13404 SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	100	246	246	
662	3	746	1399	emb X13404 SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	99	653	654	
662	1	956	480	gb U13177	[S. aureus sigma factor (p1aC) gene, complete cds	100	136	477	
665	1	1182	592	gb U05000	Staphylococcus aureus type-I signal peptidase SpsA (spsA) gene, and type-I signal peptidase SpsB (spsB) gene, complete cds	98	534	591	
665	2	1716	1153	gb U65000	Staphylococcus aureus type-I signal peptidase SpsA (spsA) gene, and type-I signal peptidase SpsB (spsB) gene, complete cds	96	564	564	
697	1	3	527	gb H6177	S. aureus sigma factor (p1aC) gene, complete cds	100	195	525	
697	2	485	1784	gb H6177	S. aureus sigma factor (p1aC) gene, complete cds	97	280	100	

TABLE 1

S. aureus - Coding regions containing known sequences

Contig	ORF	Start	Stop	match	match gene name	percent	ISP at	Out nt
ID	ID	int	int	accession		Ident	length	length
710	1	15	503	[gb DR6240 DR62	[staphylococcus aureus gene for unknown function and dlt operon dltA, dldC, dltC and dltD genes, complete cds	99	217	489
723	1	26	205	[gb HR0252	[staphylococcus aureus norA199 gene (which mediates active efflux of fluoroquinolones), complete cds			
741	1	1736	1197	[gb DR3951 STAL	[staphylococcus aureus DNA for lukM component, lukF-PV like component, complete cds	81	522	540
752	1	1	636	[emb Y00336 SASP	[staphylococcus aureus V8 serine protease gene	99	618	616
752	1	588	956	[emb Y00336 SASP	[staphylococcus aureus V8 serine protease gene	99	140	369
756	1	1308	709	[emb X01648 SATO	[staphylococcus aureus (Wood 46) gene for alpha-toxin	98	567	600
777	1	1582	950	[emb Z49205 SA42	[S. aureus partial cod gene for superoxide dismutase	99	429	633
780	1	1111	557	[gb D20503	[staphylococcus aureus MHC class-II analog gene, complete cds	86	550	555
784	1	73	687	[gb D163229	[staphylococcus aureus novel antigen gene, complete cds	99	568	615
797	1	182	544	[gb D14711 STAH	[staphylococcus aureus HSP10 and HSP60 genes	98	163	263
798	1	532	302	[emb X56434 SRPD	[S. aureus pnbB, pnbC and pnbD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	95	196	211
823	1	3	467	[gb S77055	[recF cluster: dnaA replication assembly protein...grB/DNA gyrase beta subunit] [staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	99	156	465
848	1	348	175	[gb J25488	[staphylococcus aureus gyrase-like protein alpha and beta subunit (grA and grB) genes, complete cds	99	174	174
849	2	476	318	[gb J25488	[staphylococcus aureus gyrase-like protein alpha and beta subunit (grA and grB) genes, complete cds	100	131	159
866	1	792	397	[emb X61172 SARP	[S. aureus rplB, rplC, rpsB, rpsD, rpsI and rpsJ genes for ribosomal protein L7/L12, DNA-directed RNA polymerase beta L beta' chain]	99	395	396
883	1	1	285	[gb D90119 STAN	[S. aureus nota gene	99	131	285
884	1	606	334	[emb X51543 SAG	[S. aureus agrA, agrB and hid genes	98	265	273
914	2	716	522	[emb X5241 SAG	[S. aureus agrA, agrB and hid genes	100	195	195
912	2	517	681	[emb Z20588 SAST	[S. aureus (RNA220) genes for potential AIC transporter and potential membrane spanning protein	99	163	165
917	1	2	265	[gb H4724	[S. aureus tagtose 6-phosphate isomerase gene, complete cds	99	247	264
917	2	218	196	[gb H4724	[S. aureus tagtose 6-phosphate isomerase gene, complete cds	95	147	159
918	1	2426	1215	[emb X92205 SAPT	[S. aureus pstA and ptsI genes	99	1212	1212

TABLE 1

S. aureus - Cont'd regions containing known sequences

contig	ORF ID	start (nt)	Stop (nt)	match accession	match gene name	percent ISP nt ident	ORF nt length
967	1	1	411	[dbj WP0119 STAN]	[<i>S. aureus</i> nora gene]	97	395
991	1	672	1337	[emb KS5543 SAAG]	[<i>S. aureus</i> agrA, agrB and hid genes]	99	336
1000	1	1117	845	[gb U11017]	[<i>Staphylococcus aureus</i> methicillin-resistance protein (mecc) gene and unknown ORF, complete cds]	78	150
1001	1	498	265	[dbj DR8240 DB62]	[<i>Staphylococcus aureus</i> gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds]	99	234
1010	1	1	285	[gb U21221]	[<i>Staphylococcus aureus</i> hyaluronate lyase (hyalA) gene, complete cds]	99	224
1046	1	656	330	[emb X72700 SAPV]	[<i>S. aureus</i> gene for S and F components of Panton-Valentine leukocidins]	85	205
1050	1	480	1286	[emb KS5434 SAPD]	[<i>S. aureus</i> pdnb, pdnc and pdhd genes for pyruvate decarboxylase, dihydrolipamide acetyltransferase and dihydrolipamide dehydrogenase]	99	180
1073	1	1176	589	[dbj K02885]	[<i>S. aureus</i> lata strain RN4501 transposon Tn554 insertion site]	100	131
1079	1	3	230	[dbj DR8240 DB862]	[<i>Staphylococcus aureus</i> gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds]	99	228
1079	2	218	484	[dbj DR8240 DB62]	[<i>Staphylococcus aureus</i> gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds]	100	267
1079	1	460	645	[dbj DR8240 DRG2]	[<i>Staphylococcus aureus</i> gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds]	100	186
1092	1	289	146	[emb KS5434 SAPD]	[<i>S. aureus</i> pdnb, pdnc and pdhd genes for pyruvate decarboxylase, dihydrolipamide acetyltransferase and dihydrolipamide dehydrogenase]	98	124
1143	1	1	243	[gb M63177]	[<i>S. aureus</i> alpha factor (plac) gene, complete cds]	99	241
1157	1	1	136	[emb Z48003 SA09]	[<i>S. aureus</i> gene for DNA polymerase III]	97	127
1169	1	720	361	[emb S74031]	[nora NORA (TSP794) (<i>Staphylococcus aureus</i> , NCTC 8325, Insertion, 1820 nt)]	99	360
1190	1	2	283	[dbj M21854]	[<i>S. aureus</i> agr gene encoding an accessory gene regulator protein, complete cds]	100	282
1225	1	2	1127	[emb X52543 SAAG]	[<i>S. aureus</i> agrA, agrB and hid genes]	100	240
1243	1	2	529	[dbj DR8240 DB62]	[<i>Staphylococcus aureus</i> gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds]	99	495
1244	1	1	210	[gb U574011]	[norA NORA (TSP794) (<i>Staphylococcus aureus</i> , NCTC 8325, Insertion, 1820 nt)]	100	210
1301	1	1	41	[emb X76490 SGNL]	[<i>S. aureus</i> (U5701) glnA and glnB genes]	99	299

TABLE I

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	Percent Ident	USP nt length	ORF nt length
1315	1	18	326	[gb X64172 SARP	[S. aureus] rp01, rp01rif1 and rp0c genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA Polymerase beta, L	98	277	309
1519	1	2	175	[dbj D28879 STAP	[Staphylococcus aureus] gene for penicillin-binding protein 1, complete cds	98	139	174
1663	1	1346	675	[dbj D86240 D662	[Staphylococcus aureus] gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	98	672	672
1797	1	644	324	[gb U07374	[Staphylococcus aureus] type A capsule genes, capA, capB, capBC, capD, capE, capFG, capGH, capIJ, capKL, capMH, capPN, capQO, complete cds	99	321	321
1857	1	1	192	[gb U90356	[Staphylococcus aureus] alpha-hemolysin gene, 3' end	98	192	192
1923	1	2	181	[gb X17601 SFAF	[S. aureus] factor essential for expression of methicillin resistance (femA) gene, complete cds, and tRNA gene, 3' end	100	180	180
1957	1	2	346	[gb U60891	[Staphylococcus aureus] novel antigen gene, complete cds	99	345	345
1980	1	1	402	[dbj D86240 D662	[Staphylococcus aureus] gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	100	402	402
2100	1	414	208	[gb H63177	[S. aureus] sigma factor (σ ₇₀) gene, complete cds	99	207	207
2199	1	1	402	[gb U66664	[Staphylococcus aureus] DNA fragment with class II promoter activity	99	131	402
2517	1	308	156	[gb X17601 SFAF	[S. aureus] factor essential for expression of methicillin resistance (femA) gene, complete cds, and tRNA gene, 3' end	99	153	153
2891	1	2	400	[gb J25426	[Staphylococcus aureus] penicillin-binding protein 2 (pbp2) gene, complete cds	99	399	399
2950	1	778	398	[dbj D06901 STAN	[Staphylococcus aureus] genes for ORF37; HSP20; HSP40; ORF35, complete cds	100	358	381
2971	1	3	398	[gb U51322	[Staphylococcus aureus] succinylbenzoic acid CoA ligase (menC) and α-succinylbenzoic acid synthetase (menC) genes, complete cds	97	272	396
2985	1	618	328	[gb U11979	[Staphylococcus aureus] chorismate synthase (aroC) and nucleoside diphosphate kinase (nuk) genes, complete cds, dehydroquinase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gaecC) genes, partial cds	98	250	291
3006	1	2110	1784	[gb U11791	[Staphylococcus aureus] methicillin-resistant ATCC 33952 clone RBN30 16S-21S rRNA spacer region	87	82	387
3008	1	474	238	[dbj D06901 STAN	[Staphylococcus aureus] genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	88	178	237
3008	2	451	281	[dbj D06901 STAN	[Staphylococcus aureus] genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	97	320	171

TABLE I

S. aureus - Coding regions containing known sequences

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	Percent ident	HSP length	ORF nt. length
3011	1	793	198	[emb X62991 SAMP]	[S. aureus] fmbB gene for fibronectin binding protein B	93	72	396
3019	1	2	235	[gb J03479]	[S. aureus] enzyme III-lac (lacE), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	234	234
3023	1	81	233	[gb J06451]	[Staphylococcus aureus] proline permease homolog (rpocP) gene, complete cds	87	100	153
3029	1	90	287	[gb J05133]	[Staphylococcus aureus] phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	100	135	198
3039	1	18	164	[gb J05133]	[Staphylococcus aureus] phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	97	135	147
3039	1	70	327	[gb J05133]	[Staphylococcus aureus] phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	77	183	258
3056	1	3	215	[emb X61772 SARP]	[S. aureus] rplB, orf202, rpoc genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	213	213
3059	1	1	261	[abj 030690 STAN]	[Staphylococcus aureus] genes for ORF17, HSP20, HSP40, ORF35, complete cds	98	234	261
3073	1	27	204	[gb J06451]	[Staphylococcus aureus] proline permease homolog (rpocP) gene, complete cds	99	229	258
3074	1	2	397	[emb X64172 SARP]	[S. aureus] rplC, orf202, rpoc genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	96	250	296
3088	1	3	1239	[dbj 066722 D0667]	[Staphylococcus aureus] DNA for DNA polymerase III, complete cds	95	215	237
3097	1	444	1 244	[emb J24803 SAMP]	[S. aureus] gene for DNA polymerase III	97	160	201
3102	1	307	155	[gb J03479]	[S. aureus] enzyme III-lac (lacE), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	142	153
3121	1	568	398	[emb X58334 SAMP]	[S. aureus] pdha, pdhb and pdhd genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	100	88	171
3125	1	463	233	[emb J89233 SAMP]	[S. aureus] DNA for rpoc gene	98	192	211
3133	1	2	175	[emb J218652 SACF]	[S. aureus] gene for clumping factor	96	154	174
3140	1	420	211	[dbj 010489 STAC]	[Staphylococcus aureus] genes for DNA gyrase A and B, complete cds	89	197	210
3176	1	1	378	[emb J58334 SAMP]	[S. aureus] pdha, pdhb and pdhd genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	96	91	378
3192	1	420	211	[gb J03479]	[S. aureus] enzyme III-lac (lacE), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	72	210
3210	1	3	143	[gb J076714]	[Staphylococcus aureus] peptidoglycan hydrolase gene, complete cds	96	141	141

TABLE 1

S. aureus - Coding regions containing known sequences

Config	ORF	Start	Stop	match	match gene name	Percent	ISPF NC	ORF nt	
ID	ID	(nt)	(nt)	accession		Ident	length	length	
3332	3	2106	1282	gb L14017	Staphylococcus aureus methicillin-resistance protein (mact) gene and unknown ORF, complete cds	71	257	825	
3538	1	1	2	374	emb XK923 SACP S. aureus DNA for rpoC gene	99	316	393	
3543	1	192	634	gb L11530	Staphylococcus aureus transfer RNA sequence with two rRNAs	99	102	243	
3555	1	637	320	emb Z18652 SACP S. aureus gene for clumping factor		99	307	318	
3559	1	1	3	182	emb XK17679 SACP Staphylococcus aureus coe gene for coagulase	100	141	180	
3559	2	1	95	113	emb XK17679 SACO Staphylococcus aureus coa gene for coagulase	98	174	219	
3563	1	278	141	gb L05573	Staphylococcus aureus proC; 'protein diacylglycerol transferase (lgt) gene, complete cds	100	79	138	
3563	2	527	163	gb L05573	Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds	98	162	165	
3566	1	1	3	422	emb X16157 SACST Staphylococcus aureus gene for staphylococcalase	98	175	420	
3588	1	2	262	gb L03098	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	253	261	
3593	1	1	156	gb J31479	S. aureus enoyl III-lac lacF1, enzyme II-lac (lacF), and phospho-beta-galactosidase (lacG) genes, complete cds	99	345	346	
3600	1	758	381	emb J21852 SACF S. aureus gene for clumping factor		72	346	378	
3602	1	788	396	emb J21852 SACF S. aureus gene for clumping factor		98	319	393	
3656	1	1013	528	emb J21852 SACF S. aureus gene for clumping factor		84	403	406	
3682	1	3	236	emb J264172 SACP S. aureus rplL, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta 6	100	231	234		
3682	2	224	415	emb J264172 SARP S. aureus rplL, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta 4	100	112	132		
3693	1	758	1	423	emb J262992 SAFN S. aureus fimb gene for fibronectin binding protein B				
3702	1	1	591	154	gb L11510	Staphylococcus aureus transfer RNA sequence with two rRNAs	54	61	210
3725	1	928	463	emb J21852 SACF S. aureus gene for clumping factor		71	367	462	
3761	1	809	450	gb L14017	Staphylococcus aureus methicillin-resistance protein (mact) gene and unknown ORF, complete cds	85	333	360	
3767	1	1	402	emb J264172 SACP S. aureus rplL, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta 6	98	387	402		

TABLE 1

.../11/93 - Coding regions containing known sequences

Contig	ORF	Start	Stop	match	match gene name	percent ISF nt	ORF nt
ID	ID	ID	ID	accession	(r)	ident	length
3775	1	2	286	emb X61721 SARP	[S. aureus rplL, rpsB20, rpsC genes for ribosomal protein L7/L12 hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	227
							285
3786	1	456	229	gb J0489 STAG	[Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	204
							228
3786	2	542	366	gb J0489 STAG	[Staphylococcus aureus genes for DNA gyrase A and B, complete cds	95	123
							177
3798	1	3	251	emb X17679 SACO	[Staphylococcus aureus coa gene for coagulase	99	249
							259
3813	1	793	398	gb J04151	[S. aureus fibronectin-binding protein (fnba) mRNA, complete cds	98	396
							396
3819	1	184	402	emb X64425 SAC23	[S. aureus gene for 2S rRNA	99	161
							219
3844	1	932	468	gb J048226	[Staphylococcus aureus elastin binding protein (ebps) gene, complete cds	87	204
							465
3845	1	1	381	emb X58434 SARD	[S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroxyacetone phosphate acetyltransferase and dihydroxyacetone dihydrogenase	94	356
							351
3856	1	798	400	gb J14017	[Staphylococcus aureus methicillin-resistance protein (mecA) gene and unknown ORF, complete cds	76	192
							359
3859	1	1049	573	emb Z18851 SICP	[S. aureus gene for clumping factor	85	347
							477
3871	1	650	327	gb N16714	[Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	299
							324
3876	1	2	253	gb J04481 STAG	[Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	217
							252
3877	1	572	288	gb J03479	[S. aureus enzyme 11-lac (lacE), enzyme 11-lac (lacZ) and phospho-beta-galactosidase (lacZ) genes, complete cds	97	209
							265
3878	1	1	237	emb X58434 SARD	[S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroxyacetone phosphate acetyltransferase and dihydroxyacetone dihydrogenase	96	155
							237
3888	1	3	173	emb X16457 SAST	[Staphylococcus aureus gene for staphylococulase	98	121
							171
3893	1	1	183	emb X89233 SARP	[S. aureus DNA for rpsC gene	100	170
							183
3893	2	181	357	emb J89233 SARP	[S. aureus DNA for rpsC gene	98	79
							177
3894	1	3	485	emb X64172 SARP	[S. aureus rplL, rpsB20, rpsC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	450
							493
3895	1	836	420	gb J04151	[S. aureus fibronectin-binding protein (fnba) mRNA, complete cds	99	411
							417
3905	1	48	239	gb L05004	[Staphylococcus aureus dehydroquinate synthase (tarob) gene, 3' end cds; 3'-phosphoshikimate-1-carboxyvinyltransferase (tarob) gene, complete cds; ORF, complete cds	100	159
							192
3905	2	188	400	gb L05004	[Staphylococcus aureus dehydroquinate synthase (tarob) gene, 3' end cds; 3'-phosphoshikimate-1-carboxyvinyltransferase (tarob) gene, complete cds; ORF, complete cds	97	86
							213

TABLE 1

S. aureus - Coding regions containing known sequences

config	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent HSP nt	ORF nt
		(nt)	(nt)	[...]		ident	length
3910	1	3	359	[emb]X58434 SAPD	[S. aureus] <i>pdhA, pdhC and pdhD genes for pyruvate decarboxylase, dihydroxyacetone acetyltransferase and dihydroxyacetone dehydrogenase</i>	99	278
3915	1	1	330	[gb L10171]	[Staphylococcus aureus] <i>methicillin-resistance protein (mecR) gene and unknown ORF, complete cds</i>	75	175
3964	1	691	347	[emb]Z48003 SADN	[S. aureus] <i>gene for DNA polymerase III</i>	100	295
4007	1	199	390	[emb]X16157 SAST	[Staphylococcus aureus] <i>gene for staphylococulase</i>	98	163
4016	1	3	371	[dbj]D0489 STAG	[Staphylococcus aureus] <i>genes for DNA gyrase A and B, complete cds</i>	99	339
4046	1	692	348	[emb]Z11832 SACF	[S. aureus] <i>gene for clumping factor</i>	87	221
4060	1	1	375	[emb]Z88832 SACF	[S. aureus] <i>gene for clumping factor</i>	1	96
4061	1	860	432	[emb]Z88003 SADN	[S. aureus] <i>gene for DNA polymerase III</i>	99	429
4062	1	606	304	[gb L10171]	[Staphylococcus aureus] <i>methicillin-resistance protein (mecR) gene and unknown ORF, complete cds</i>	75	198
4085	1	58	402	[gb L17861]	[Staphylococcus aureus] <i>methicillin-resistant ATCC 33952 clone RUUV2 16S-23S rRNA spacer region</i>	98	127
4088	1	2	301	[gb L41098]	[Transposon] <i>Tn5104 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) URA</i>	99	227
4093	1	2	277	[emb]X28414 SAPD	[S. aureus] <i>pdhA, pdhC and pdhD genes for pyruvate decarboxylase, dihydroxyacetone acetyltransferase and dihydroxyacetone dehydrogenase</i>	99	276
4097	1	1	402	[emb]Z11832 SACF	[S. aureus] <i>gene for clumping factor</i>	74	307
4116	1	22	402	[gb L03004]	[Staphylococcus aureus] <i>dehydroquinate synthase (aroD) gene, 3' and cds; 3'-phosphoshikimate-1-carboxyvinyltransferase (aroK) gene, complete cds; ORF3, complete cds</i>	98	157
4125	1	240	101	[gb L07174]	[Staphylococcus aureus] <i>type 8 capsule genes, capA, capB, capC, capD, capE, capF, capG, capH, capI, capK, capM, capN, capO, capP, capQ, capR, capS, capT, capU, capV, capW, capX, capY, capZ, capAA, capAB, capAC, capAD, capAE, capAF, capAG, capAH, capAI, capAJ, capAK, capAL, capAM, capAN, capAO, capAP, complete cds</i>	100	86
4149	1	35	247	[gb J0151]	[S. aureus] <i>fibronectin-binding protein (fnbA) mRNA, complete cds</i>	99	200
4151	1	629	366	[gb L10171]	[Staphylococcus aureus] <i>methicillin-resistance protein (mecR) gene and unknown ORF, complete cds</i>	87	150
4154	1	754	398	[emb]X24172 SARP	[S. aureus] <i>rpB (rpf) and rpC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta 4 beta-chains</i>	99	297
4179	1	1	294	[emb]X24172 SARP	[S. aureus] <i>rpB (rpf) and rpC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta 4 beta-chains</i>	98	240

TABLE 1

S. aureus - Coding regions containing known sequences

Contig	ORF	Start	Stop	match gene name	percent ISP nt	ORF nt
ID	TU	(nt)	(nt)	match	Ident	length
4203	1	1	255	emb X89233 SARP <i>S. aureus</i> DNA for rpoC gene	99	239
4206	1	1	303	emb 218852 SACP <i>S. aureus</i> gene for clumping factor	100	236
4206	2	195	344	emb 218852 SACP <i>S. aureus</i> gene for clumping factor	95	65
4208	1	108	314	emb X86434 SARD <i>S. aureus</i> pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	89	76
4216	1	656	220	emb X86434 SARD <i>S. aureus</i> pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	126
4226	1	594	298	gb 11530 <i>Staphylococcus aureus</i> transfer RNA sequence with two rRNAs	97	132
4260	1	216	383	gb 11784 <i>Staphylococcus aureus</i> methicillin-resistant ATCC 33952 clone RNNAO 16S-23S rRNA spacer region	83	141
4272	1	355	179	emb 24003 SAON <i>S. aureus</i> gene for DNA polymerase III	100	164
4276	1	4	177	emb X16437 SAST <i>Staphylococcus aureus</i> gene for staphylocoagulase	99	150
4277	1	1	270	emb X64172 SARP <i>S. aureus</i> rplU, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta 6	99	265
4282	1	691	377	emb X64172 SARP <i>S. aureus</i> rplU, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta 6	98	282
4291	1	179	191	emb X64172 SARP <i>S. aureus</i> rplU, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta 6	99	183
4395	1	3	329	emb X16437 SAST <i>Staphylococcus aureus</i> gene for staphylocoagulase	94	144
4313	1	435	280	gb 11530 <i>Staphylococcus aureus</i> transfer RNA sequence with two rRNAs	100	94
4315	1	3	185	gb 103479 <i>S. aureus</i> enzyme III-lac (lacZ), and phospho-beta-galactosidase (lacG) genes, complete cds	100	158
4315	2	101	310	gb 103479 <i>S. aureus</i> enzyme III-lac (lacZ), enzyme III-lac (lacZ), and phospho-beta-galactosidase (lacG) genes, complete cds	98	75
4327	1	1	294	gb 141098 Transposon Tn504 and insertion sequence IS1181 and IS1182 (from <i>Staphylococcus aureus</i>) DNA	98	294
4360	1	603	319	gb 103910 <i>Staphylococcus aureus</i> ATCC 25923 16S rRNA gene, partial sequence	100	116
4364	1	3	146	emb X64172 SARP <i>S. aureus</i> rplU, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta 6	95	140
4368	1	167	310	emb X22932 SAFN <i>S. aureus</i> fibronectin binding protein B	73	119
						144

TABLE 1

S. aurous - Coding regions containing known sequences

5
10
15
20
25
30
35
40
45
50

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ISP nt ident	ORF nt length
4401	1	2	113	emb X6992 SATN	S. aureus fnbB gene for fibronectin binding protein B	97	243
4421	1	36	281	dbj 013572 STAA2	Staphylococcus aureus rRNA gene for 23S ribosomal RNA	100	112
4426	1	3	293	emb 218852 SACF	S. aureus gene for clumping factor	85	185
4428	1	493	248	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA Polymerase beta & beta' chains	100	139
4462	1	2	271	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA Polymerase beta & beta' chains	99	270
4466	1	1	240	emb 218852 SACF	S. aureus gene for clumping factor	99	231
4469	1	1	312	gb J03479	S. aureus enzyme III-lac (lacG), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	99	265
4485	1	3	263	gb LA3098	Transposon Tn404 and insertion sequences IS1161 and IS1162 (from Staphylococcus aureus) DNA	98	259
4492	1	74	400	gb MB8227	Staphylococcus aureus DNA gyrase B subunit (gyrB) RecF homologue (recf) and DNA gyrase A subunit (gyrA) gene, complete cds	85	104
4497	1	515	269	emb 218852 SACF	S. aureus gene for clumping factor	99	213
4529	1	2	172	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA Polymerase beta & beta' chains	100	151
4547	1	1	300	emb X6992 SATN	S. aureus fnbB gene for fibronectin binding protein B	100	157
4554	1	218	160	emb 218852 SACF	S. aureus gene for clumping factor	84	126
4565	1	9	237	emb 218852 SACF	S. aureus gene for clumping factor	84	213
4569	1	79	222	emb 218852 SACF	S. aureus gene for clumping factor	98	127
4608	1	22	216	emb X56434 SAPD	S. aureus pfbB, pfbC and pfbD genes for pyruvate decarboxylase, dihydroxyacetone acetyltransferase and dihydroxyacetone dehydrogenase	92	168
4614	1	464	234	emb 218852 SACF	S. aureus gene for clumping factor	86	169
4623	1	105	302	gb J04151	S. aureus fibronectin-binding protein (fiba) mRNA, complete cds	99	152
4632	1	18	206	gb J03479	S. aureus enzyme III-lac (lacE), enzyme II-lac (lacZ), and phospho-beta-galactosidase (lacG) genes, complete cds	98	183
4646	1	1	222	emb 218852 SACF	S. aureus gene for clumping factor	84	100
4687	1	2	166	gb J04151	S. aureus fibronectin-binding protein (fiba) mRNA, complete cds	98	156

TABLE 1

S. aureus - Coding regions containing known sequences

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent HSP nt Ident	ORF nt length
4695	1	313	158	gi 114017	[Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds]	75	155
4703	1	1	153	emb X55843 SAPD	[S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase]	98	103

TABLE 1

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
20	6	5089	4679	91 1511839	ORF1 [Staphylococcus bacteriophage phi 11]	100	100	411
149	3	2022	1577	91 1B9703 B497	int gene activator RNA - bacteriophage phi 11	100	100	456
149	5	2109	1912	91 166161	Bacteriophage phi-11 Int gene activator [Staphylococcus bacteriophage phi 11]	100	100	198
149	2	558	409	91 166159	integrase (int) [Staphylococcus bacteriophage phi 11]	100	100	150
396	1	1372	707	91 166159	Integrase (int) [Staphylococcus bacteriophage phi 11]	100	99	666
398	2	783	1001	91 1455128	[excisionase (xis) [Staphylococcus bacteriophage phi 11]	100	100	219
502	4	13914	1744	91 1204912	[H. influenzae predicted coding region H10660] [Haemophilus influenzae]	100	71	171
849	1	2	262	91 1373002	[polyprotein (Bean common mosaic virus)]	100	46	261
1149	1	277	140	91 143359	protein synthesis initiation factor 2 (infB) [Bacillus subtilis] 91 69319	100	82	138
2880	1	21	308	91 1662933	[infB gene product (Bacillus subtilis)]	100	98	288
3045	1	428	216	91 1354211	[protein kinase C inhibitor-1 (Homo sapiens)]	100	100	213
4168	2	571	398	91 1354211	[PET12-like protein (Bacillus subtilis)]	100	100	174
331	1	2	247	91 1426673	[nuoC gene product (Staphylococcus carnosus)]	98	95	246
207	2	1272	1463	91 1460259	[enolase (Bacillus subtilis)]	97	90	192
231	1	2	375	91 1581638	[lal protein (Staphylococcus carnosus)]	97	53	456
366	1	39	215	91 166161	Bacteriophage phi-11 Int gene activator [Staphylococcus bacteriophage phi 11]	97	95	177
660	3	718	936	91 1426173	[nuoC gene product (Staphylococcus carnosus)]	97	97	219
1578	1	284	144	91 11339250	[large subunit of NADH:ubiquinol oxidoreductase (I) [Plectonema incognitus]]	97	79	141
157	1	321	518	91 1023726	[Unknown (Staphylococcus haemolyticus)]	96	88	198
205	133	16670	16147	91 11465302	[S10 (Bacillus subtilis)]	96	91	324
3919	1	4R	401	91 1971784	[Clp-like ATP-dependent protease binding subunit [Bos taurus]]	96	81	354
4133	1	830	417	91 1022726	[Unknown (Staphylococcus haemolyticus)]	96	84	414
4168	1	708	355	91 1354211	[PET12-like protein (Bacillus subtilis)]	96	95	354
4207	1	312	157	91 6202031	similar to trimethyllysine DH [Hycoplasma capricolum] p1r1S4950 S49950 probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolum (SEC) (fragment)	96	86	156

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start ID	Stop ID	match accession	match gene name	% sim	% ident	length (nt)
4327	1	152	331	g1 1971784	[Clp-like ATP-dependent protease binding subunit [Bos taurus]	96	81	180
4416	1	570	286	g1 1022726	[Unknown (Staphylococcus haemolyticus)]	96	84	285
22	1	858	430	g1 1511070	[ureC (Staphylococcus xylosus)]	95	88	429
22	7	4362	4036	g1 1581787	[urease gamma subunit (Staphylococcus xylosus)]	95	79	327
62	6	8794	9114	lpir JCG008 JAC00	[ribosomal protein S7 - Bacillus stearothermophilus]	95	83	321
154	9	920	7838	g1 1354211	[PEP12-like protein (Bacillus subtilis)]	95	92	1443
186	3	2798	2055	g1 1514656	[serine O-acetyltransferase (Staphylococcus xylosus)]	95	87	744
205	5	4406	4014	g1 142462	[ribosomal protein S11 (Bacillus subtilis)]	95	85	393
205	7	5017	4793	g1 142659	[initiation factor 1 (Bacillus subtilis)]	95	84	225
205	21	11365	10991	g1 1044974	[ribosomal protein L14 (Bacillus subtilis)]	95	93	375
259	5	7288	6644	lpir P47995 YSEA_	[HYPOTHETICAL PROTEIN IN SECA S REGION (OF1)] (FRAGMENT)	95	85	645
302.	3	795	1097	g1 40186	[homologous to E. coli ribosomal protein L27 (Bacillus subtilis) l1 143592 L27 ribosomal protein (Bacillus subtilis) l1 C1885 C21895 ribosomal protein (Bacillus subtilis) l1 PUS657 R127_DACSU 50S RIBOSOMAL PROTEIN L27 (BL30) (BL241). 1 40175 L24 gene prod]	95	89	303
310	1	579	1523	g1 1177684	[chorismate mutase (Staphylococcus xylosus)]	95	92	945
414	1	2	163	lpir C483996 C483	[ribosomal protein L34 - Bacillus stearothermophilus]	95	90	162
4165	2	125	277	g1 1276841	[glutamate synthase (GOGAT) (Porphyra purpurea)]	95	86	153
22	2	1028	723	g1 1511069	[ureF (Staphylococcus xylosus)]	94	91	306
22	5	5016	3310	g1 410516	[urease alpha subunit (Staphylococcus xylosus)]	94	85	1737
60	4	815	1372	g1 1666116	[glucose kinase (Staphylococcus xylosus)]	94	87	558
205	18	110012	9536	g1 1044978	[ribosomal protein S8 (Bacillus subtilis)]	94	78	477
326	4	3378	2542	g1 557492	[dihydroxynaphthalic acid (DNA) synthetase (Bacillus subtilis) l1 143186 (Bacillus subtilis)]	94	65	637
414	3	737	955	g1 1467396	[thiophen and furan oxidation (Bacillus subtilis)]	94	77	219
426	3	2260	1823	g1 1263908	[putative (Staphylococcus epidermidis)]	94	87	438
524	1	2	335	g1 1633650	[enzyme II (mannitol) (Staphylococcus cereus)]	94	84	354
1017	1	2	229	g1 149435	[putative (lactococcus lactis)]	94	73	228
2098	1	330	184	g1 141952	[lpa-2d gene product (Bacillus subtilis)]	94	50	147

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start	Stop	match accession	match gene name	sim	ident	length (nt)
3232	1	630	316	[gi 1022755	[unknown (Staphylococcus haemolyticus)]	94	84	315
42	5	2089	2359	[pir BA8396 B483	[ribosomal protein L33 - <i>Bacillus stearothermophilus</i>	93	81	171
101	2	1745	1383	[gi 1155345	[secretin efflux pump protein (Plasmid pK267)]	93	82	363
205	124	12227	11865	[esp P16577 P016-	[50S RIBOSOMAL PROTEIN L16]	93	83	363
259	4	8391	5873	[gi 1499335	[leucK protein (Staphylococcus carnosus)]	93	85	2619
275	1	2226	1114	[gi 613650	[enzyme II (mannitol) (Staphylococcus carnosus)]	93	86	1113
444	6	6207	5773	[gi 1022726	[unknown (Staphylococcus haemolyticus)]	93	81	435
491	1	152	622	[gi 146912	[ribosomal protein L13 (Staphylococcus carnosus)]	93	88	471
607	6	1674	2033	[gi 1022726	[unknown (Staphylococcus haemolyticus)]	93	83	360
653	1	973	488	[gi 1580950	[translation initiation factor IF3 (AA 1-122) (Bacillus teatothermophilus)]	93	77	466
1864	1	3	194	[gi 106553	[ribosomal protein small subunit (Homo sapiens)]	93	93	192
2397	1	20	300	[gi 143390	[carboxy] phosphate synthetase (Bacillus subtilis)]	93	82	273
3232	2	907	596	[gi 1022725	[unknown (Staphylococcus haemolyticus)]	93	84	312
3761	2	794	621	[gi 1022725	[unknown (Staphylococcus haemolyticus)]	93	88	174
16	1	3	374	[gi 142781	[putative cytoplasmic protein, putative (Bacillus subtilis) sp P37954 UNKB_BACSU_EXCINUCLEASE_A6C SUBUNIT_B (DNA PROTEIN) FRAGMENT]	92	83	372
31	7	5915	6124	[gi 1136430	[K11A0185 protein (Homo sapiens)]	92	46	210
56	19	126483	127391	[gi 1467101	[unknown (Bacillus subtilis)]	92	80	909
69	6	5882	6130	[gi 1530200	[crophoblastin (Ovis aries)]	92	53	249
135	3	2568	2038	[gi 1022725	[unknown (Staphylococcus haemolyticus)]	92	80	521
171	3	2760	2362	[gi 5157475	[D-amino acid transaminase (Staphylococcus haemolyticus)]	92	86	399
205	122	7495	6962	[gi 1491889	[secY gene product (Staphylococcus carnosus)]	92	85	534
205	119	10012	110255	[gi 1044976	[ribosomal protein L5 (Bacillus subtilis)]	92	82	558
219	1	710	1357	[gi 1103812	[Yav (Bacillus subtilis)]	92	88	354
344	3	1575	1805	[gi 1405474	[cpc protein (Bacillus cereus)]	92	85	231
699	1	20	361	[gi 413939	[lpa-75d gene product (Bacillus subtilis)]	92	81	342
1343	1	2	160	[pir A4534 A454	[ribosomal protein L19 - <i>Bacillus stearothermophilus</i>]	92	84	159

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start	Stop	match accession	match gene name	sim	ident	length (nt)
ID	ID	(nt)	(nt)					
1958	1	524	264	gi 407908	lfrtrc [Staphylococcus xylosus]	92	80	261
3578	2	716	286	gi 1339950	large subunit of NADH-dependent glutamate synthase (Pleotomace boryanum)	92	78	333
1585	1	644	324	gi 1339950	large subunit of NADH-dependent glutamate synthase (Pleotomace boryanum)	92	81	321
3610	1	4	402	gi 1022726	lunknow [Staphylococcus haemolyticus]	92	81	399
4342	1	14	178	gi 450688	lhdB gene product (Escherichia coli) pir S1847 S1817 haOH	92	74	165
					protein - Escherichia coli pir S0921 S09629 hypothetical protein A -			
					Escherichia coli (sub 40-520)			
4446	1	358	182	gi 1022725	lunknow [Staphylococcus haemolyticus]	92	82	177
4519	1	462	232	gi 11022726	lunknow [Staphylococcus haemolyticus]	92	80	231
4626	1	3	224	gi 1022725	lunknow [Staphylococcus haemolyticus]	92	84	222
2	4	3980	4531	gi 535349	lcodW [Bacillus subtilis]	91	74	552
28	1	2	1126	gi 11003376	hypothetical protein (Synchocystis sp.)	91	78	1125
60	5	1354	1701	gi 11226043	lafZ downstream of glucose kinase [Staphylococcus xylosus]	91	80	348
101	1	1989	1036	gi 1150728	arsenic efflux pump protein (Plasmid p1258)	91	80	954
187	2	412	1194	gi 142559	ATP synthase alpha subunit (Bacillus megaterium)	91	79	783
205	22	11579	11298	gi 140149	S17 protein (AA 1-87) [Bacillus subtilis]	91	63	282
206	7	6164	10762	gi 1072618	glcA gene product [Staphylococcus carnosus]	91	83	2079
106	2	3885	2326	gi 143012	GDP synthetase [Bacillus subtilis]	91	78	1560
106	1	5319	3826	gi 467399	IMP dehydrogenase [Bacillus subtilis]	91	79	1494
310	3	2194	3207	gi 11177685	ccpA gene product [Staphylococcus xylosus]	91	81	1014
143	4	2874	3150	gi 198974	lucose repressor [Staphylococcus xylosus]	91	82	177
480	3	1606	3042	gi 1433991	ATP synthase subunit beta [Bacillus subtilis]	91	85	1437
516	3	2026	1280	gi 143366	adenylosuccinate lyase (PUR-B) [Bacillus subtilis pir C2932 W28SDS	91	79	747
					adenylosuccinate lyase (DC 4.3-2.2) - <i>Bacillus subtilis</i>			
552	1	1064	615	gi 297874	fructose-biphosphate aldolase (Staphylococcus carnosus) pir A5981 A45943	91	79	450
					fructose-biphosphate aldolase (EC 4.1.2.13) - <i>Staphylococcus carnosus</i>			
637	1	1	1536	gi 143597	lcpB synthetase (Bacillus subtilis)	91	79	1536
659	1	21	359	gi 1385178	lunknow (Bacillus subtilis)	91	66	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	orf ID	Start (int)	Stop (int)	match accession	match gene name	% sim	% ident	length
1327	1	339	530	gi 1495558	orfX (Bacillus subtilis)	91	71	192
2515	1	466	275	gi 5151070	ureG (Staphylococcus xylosus)	91	85	192
2594	1	2	202	gi 146824	beta-cystathionease (Escherichia coli)	91	75	201
3764	1	847	425	gi 1022735	unknown (Staphylococcus haemolyticus)	91	78	423
4011	1	127	495	gi 1022726	unknown (Staphylococcus haemolyticus)	91	79	369
4227	1	1	177	gi 296664	ATPase (Lactococcus lactis)	91	66	177
42	3	615	1033	gi 520401	catalase (Haemophilus influenzae)	90	86	219
51	8	3717	4607	gi 580899	OppF gene product (Bacillus subtilis)	90	74	891
129	3	5317	4003	gi 146206	l-glutamate dihydrogenase (Bacillus subtilis)	90	76	1317
164	117	16828	16933	gi 505166185	[30S RIBOSOMAL PROTEIN S15 (Bacillus subtilis)]	90	74	306
171	5	2883	2819	gi 51517475	[D-amino acid transaminase (Staphylococcus haemolyticus)]	90	78	165
205	4	4497	3550	gi 1442463	RNA polymerase alpha-core-subunit (Bacillus subtilis)	90	76	948
205	6	4748	4410	gi 104989	[ribosomal protein S11 (Bacillus subtilis)]	90	73	339
205	10	7165	6104	gi 49189	[secY gene product (Staphylococcus carnosus)]	90	81	762
205	11	6645	6472	gi 6919	[secY gene product (Staphylococcus carnosus)]	90	78	174
205	27	13692	13345	gi 786157	[ribosomal Protein S19 (Bacillus subtilis)]	90	79	346
205	31	15858	15496	gi 1165301	[U3 (Bacillus subtilis)]	90	79	363
260	5	7023	5773	gi 1161380	[iceA (Staphylococcus epidermidis)]	90	78	1251
299	6	3378	3947	gi 467440	[phosphotriacylpyrophosphate synthetase (Bacillus subtilis) gi 10218 PRPP synthetase (AA 1-317) (Bacillus subtilis)]	90	78	570
320	2	1025	1717	gi 312443	[carbamoyl-phosphate synthase (glutamine hydrolyzing) (Bacillus alrdolyticus)]	90	75	633
330	4	1581	1769	gi 1946563	[beta-tubulin (Sporidioholobius pararoseus)]	90	80	189
369	1	1954	523	gi 5347621547	[L-serine dehydratase beta chain - Clostridium sp.]	90	77	432
557	1	3	188	gi 1511589	[H. Jannaschii predicted coding region HJ1624 (Methanococcus jannaschii)]	90	54	186
663	2	667	1200*	gi 143746	[tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Bacillus subtilis)]	90	73	534
717	1	1	261	gi 141065	[ribut (Bacillus stearothermophilus)]	90	79	261
745	4	1059	865	gi 1205433	[H. influenzae predicted coding region HII190 (Haemophilus influenzae)]	90	81	195

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start	Stop	match	match	% sim	% ident	length
10	10	Int1	Int1	accession	with gene name	Int1	Int1	Int1
1007	1	386	565	gi 143366	[<i>Bacillus subtilis</i>] prc [C29326] M2BS05 [adenylosuccinate lyase (EC 4.3.2.21) - <i>Bacillus subtilis</i>]	90	77	180
1054	1	579	331	gi 10313122	[ORF_E729] [<i>Escherichia coli</i>]	90	50	249
1156	1	117	707	gi 1477776	[CIPF [<i>Bacillus subtilis</i>]]	90	80	591
1180	1	408	205	gi 1377833	[unknown [<i>Bacillus subtilis</i>]]	90	74	204
1253	1	1	462	gi 40046	[phosphoglucomutase A (AA 1-49) [<i>Bacillus stearothermophilus</i>]] [gi 515936] [NUB5A glucose-6-phosphate isomerase (EC 5.3.1.9) A - <i>cilius</i> stearothermophilus]	90	75	462
2951	1	3	269	gi 144816	[formyltetrahydrofolate synthetase (PTHFS) (tta start codon) (EC 3.4.1) [<i>Moecilia thermocactica</i>]]	90	76	267
3140	1	127	166	gi 1070014	[protein-dependent [<i>Bacillus subtilis</i>]]	90	52	162
4594	1	3	233	gi 1871764	[Clp-like ATP-dependent protease binding subunit [<i>Bos taurus</i>]]	90	76	231
87	1	1028	1750	gi 1467327	[unknown [<i>Bacillus subtilis</i>]]	89	75	723
112	1	2	505	gi 3153791	[ATP-binding protein [<i>Streptococcus mutans</i>]]	89	77	504
118	1	120	398	gi 11030304	[YqQ [<i>Bacillus subtilis</i>]]	89	75	279
128	4	3545	3757	gi 1460257	[triose phosphate isomerase [<i>Bacillus subtilis</i>]]	89	84	213
164	12	11667	12755	gi 39954	[IP2 (aa 1-74) [<i>Bacillus stearothermophilus</i>]]	89	80	1089
205	11	786	7405	gi 216338	[ORF for L15 ribosomal protein [<i>Bacillus subtilis</i>]]	89	76	471
205	132	116152	11823	gi 11165303	[L3 [<i>Bacillus subtilis</i>]]	89	80	330
270	3	2407	2207	gi c41902 CAIG	[arasanate reductase (EC 1.1.1.1) - <i>Staphylococcus xylosus</i> plasmid pSX267]	89	81	201
395	2	157	672	gi 1520574	[glutamate racemase [<i>Staphylococcus haemolyticus</i>]]	89	80	516
494	1	3	839	gi 396259	[protease [<i>Staphylococcus epidermidis</i>]]	89	77	837
510	1	1	464	gi 40046	[phosphoglucomutase A (AA 1-49) [<i>Bacillus stearothermophilus</i>]] [gi 515936] [NUB5A glucose-6-phosphate isomerase (EC 5.3.1.9) A - <i>cilius</i> stearothermophilus]	89	74	444
615	1	2124	1210	gi 11203012	[YqU [<i>Bacillus subtilis</i>]]	89	74	915
841	1	18	341	gi 1165303	[L3 [<i>Bacillus subtilis</i>]]	89	80	324
1111	1	352	813	gi 47146	[threonine tRNA synthetase [<i>Staphylococcus intermedius</i>]]	89	70	462
1875	1	2	256	gi 1205108	[ATP-dependent protease binding subunit [<i>Haemophilus influenzae</i>]]	89	82	255
2963	1	111	367	gi 467458	[cell division protein [<i>Bacillus subtilis</i>]]	89	83	357

TABLE 2

5
10
15
20
25
30
35
40
45
50
55

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	DNF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3020	1	90	362	[gi 1239948]	hypothetical protein (Bacillus subtilis)	89	66	273
3565	1	2	400	[gi 1236635]	l-idohydroxy-acid dehydratase (Bacillus subtilis)	89	75	399
3586	1	105	214	[gi 580832]	ATP synthase subunit gamma (Bacillus subtilis)	89	82	210
3629	1	794	399	[gi 1093466]	Respiratory nitrate reductase (Bacillus subtilis)	89	78	396
3688	1	2	400	[gi 146206]	glutamate dehydrogenase (Bacillus subtilis)	89	75	399
3699	1	794	399	[gi 1339950]	large subunit of NADH-dependent glutamate synthase (Electrona boryanum)	89	75	396
4016	1	428	216	[gi 1009366]	Respiratory nitrate reductase (Bacillus subtilis)	89	71	213
4177	1	471	301	[gi 149426]	[putative] (Lactococcus lactis)	89	76	171
4436	1	601	302	[gi 1022725]	Unknown [Staphylococcus haemolyticus]	89	80	300
4635	1	120	162	[gi 1102275]	Unknown [Staphylococcus haemolyticus]	89	73	159
2	2	1130	2676	[gi 1520754]	[putative] (Bacillus subtilis)	88	76	1347
42	2	468	848	[sp P43221 LATA_]	CATALASE (EC: 1.11.1.6)	88	76	381
53	1	6189	4722	[gi 474177]	Alpha-D-1-4-glucosidase (Staphylococcus xylosus)	88	80	1668
56	16	116018	118617	[gi 1467411]	Recombination protein (Bacillus subtilis)	88	77	600
60	3	376	833	[gi 665116]	Glucose kinase (Staphylococcus xylosus)	88	77	468
70	2	1583	1245	[gi 144095]	Replication initiator protein (Listeria monocytogene)	88	74	339
82	1	111514	112719	[gi 106663 A606]	Translation elongation factor Tu - Bacillus subtilis	88	79	1206
103	1	4179	4391	[gi 167181]	Serine/threonine kinase receptor (Brassica napus)	88	77	213
114	6	1732	8332	[gi 1102276]	Unknown (Staphylococcus haemolyticus)	88	72	501
118	2	308	2031	[gi 1301804]	[YQQ] (Bacillus subtilis)	88	77	1704
141	3	657	11336	[gi 1404466]	transketolase (Bacillus subtilis)	88	72	480
148	7	5871	6116	[gi 1118002]	l-idohydroperoxide synthase (Staphylococcus haemolyticus)	88	78	246
165	3	1428	2231	[gi 100953]	(phenylalanyl-tRNA synthetase alpha subunit (Bacillus subtilis) - tRNA ligase (EC 6.1.1.20) alpha sin - Bacillus subtilis)	88	80	804
205	128	15627	141485	[gi 1163306]	[L2] (Bacillus subtilis)	88	82	843
225	1	1569	896	[gi 1301840]	[YQS] (Bacillus subtilis)	88	78	672
235	1	2	1975	[gi 152209]	[valyl-tRNA synthetase (Bacillus subtilis)]	88	76	1974

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start (nt)	Stop (nt)	match accession	match gene name	1 sim	1 ident	length (nt)
10	532	1	3	61118002	dihydropteroate synthase (Staphylococcus haemolyticus)	86	73	695
11	339	3	2060	1566	lipoicidine nucleotide phosphorylase (Bacillus subtilis)	86	73	1398
12	443	4	4225	2928	valyl-tRNA synthetase (Bacillus stearothermophilus) spf11931 [SPV_BACST]	86	73	1398
13	534	3	2504	2968	valyl-tRNA synthetase (EC 6.1.1.9) VALINE-tRNA LIGASE (VALRS).	86	78	417
14	1000	2	1824	1309	mannitol-specific enzyme-III (Staphylococcus carnosus) pir Q00881J00088	86	82	465
15	705	7	584	399	mannitol-specific enzyme-III (Staphylococcus carnosus) pir Q00881J00088	86	82	465
16	1399	1	587	324	mannitol-specific IIA component EIIA-NTU_1	86	78	516
17	1341	2	170	400	nitrite reductase (nrB) (Bacillus subtilis)	86	70	186
18	1306	1	41	214	unknown (Staphylococcus carnosus) spf11976 [SPVMA_STPACA_PTS_SYSTEM]	86	78	516
19	1316	2	183	513	lipoxygenanone-lactone (Mycofunga dicum)	86	55	264
20	2949	1	708	398	ribosomal protein L20 (AA 1-19) (Bacillus stearothermophilus) pir S05346 [SPS20_rB20] ribosomal protein L20 - Bacillus stearothermophilus	86	82	231
21	2954	1	5	169	10-acetylserine(thiol) lyase (Spinacella olaracea)	86	70	165
22	3015	1	1	138	10-acetylserine(thiol) lyase (Spinacella olaracea)	86	71	174
23	3069	1	3	152	signal recognition particle 54K chain homolog (fth - Bacillus subtilis)	86	71	151
24	3917	1	817	410	signal recognition particle 54K chain homolog (fth - Bacillus subtilis)	86	73	108
25	4199	1	606	342	101140054	86	67	138
26	4201	1	734	369	101160055	86	66	150
27	4224	1	1	336	101151938	86	77	408
28	4108	1	794	399	101116206	86	71	396
29	2	5	450	6000	1011519350	87	70	1431
30	52	8	6781	6462	1011064791	87	66	300

TABLE 2

5
10
15
20
25
30
35
40
45
50
55

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	1 sim	1 ident	length (nt)
73	3	1584	2410	gi 142992	[glycerol kinase (GPK) (EC 2.7.1.30) (Bacillus subtilis) pir B45868 B45868 glycerol kinase (EC 2.7.1.30) - Bacillus subtilis sp PIB157 GPK - PHOSPHOTRANSFERASE GLYCEROL KINASE (EC 2.7.1.30) (ATPGLYCEROL - PHOSPHOTRANSFERASE)	87	72	897
90	12	8613	9100	gi 147433	[unknown (Bacillus subtilis)]	87	62	288
124	4	4265	2988	gi 1556886	[serine hydroxymethyltransferase (Bacillus subtilis) pir S9163 S9163 serine hydroxymethyltransferase - Bacillus subtilis]	87	77	1278
124	6	4457	4012	gi 1556883	[unknown (Bacillus subtilis)]	87	66	426
148	5	3741	4559	gi 146160	[unknown (Bacillus subtilis)]	87	70	819
164	13	12210	13810	gi 119934	[IF2 (aa 1-70) (Bacillus stearothermophilus)]	87	72	1101
177	2	1104	2126	gi 1467385	[unknown (Bacillus subtilis)]	87	78	1023
199	1	1982	1158	gi 1143527	[iron-sulfur protein (Bacillus subtilis)]	87	77	825
199	2	4717	2933	gi pir A27763 A277	[succinate dehydrogenase (EC 1.3.99.1) flavoprotein - Bacillus subtilis]	87	80	1785
205	12	111762	111543	gi 1044972	[ribosomal protein S19 (Bacillus subtilis)]	87	78	240
205	125	113295	112607	gi 11165109	[S3 (Bacillus subtilis)]	87	75	659
222	1	2033	1107	gi 11177249	[recB33 gene product (Bacillus subtilis)]	87	70	927
236	3	1635	1333	gi 1146198	[ferredoxin (Bacillus subtilis)]	87	80	303
246	5	1255	2292	gi 1467373	[ribosomal protein S18 (Bacillus subtilis)]	87	77	294
260	2	4189	3422	gi 11161382	[ltsC (Staphylococcus epidermidis)]	87	72	768
320	3	1666	1239	gi 11212443	[carbamoyl-phosphate synthase (glutamine-hydrolysing) (Bacillus subtilis)]	87	60	696
340	4	1165	1383	gi 142570	[ATP synthase c subunit (Bacillus licheniformis)]	87	80	219
414	4	900	1073	gi 1467386	[thiophan and furan oxidation (Bacillus subtilis)]	87	77	174
425	2	1003	1794	gi 1046166	[lollin repressor (Hypoclosteria genitalium)]	87	69	210
446	1	1235	722	gi 105134	[acetate kinase (Bacillus subtilis)]	87	75	534
480	1	1	713	gi 1142559	[ATP synthase alpha subunit (Bacillus megaterium)]	87	79	711
481	1	2	1352	gi 006797 IRG B SOS RIBOSOMAL PROTEIN L1 (Bacillus)	87	72	351	
677	2	359	955	gi 1460911	[fructose-biphosphate aldolase (Bacillus subtilis)]	87	78	597
677	3	934	1284	gi 1460911	[fructose-biphosphate aldolase (Bacillus subtilis)]	87	78	351
876	1	3	452	gi 11146247	[lasparginyl-tRNA synthetase (Bacillus subtilis)]	87	79	450

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	DNF ID	Start (int)	Stop (int)	match accession	match gene name	% sim	% ident	length (int)
1376	1	426	214	911065555	[F16H6.4 gene product (Cenorhabditis elegans)]	87	75	213
2206	1	3	374	911205098	[restriction endonuclease (Bacteriophage 154e)]	87	72	372
2398	1	3	290	911508979	[GTP-binding protein (Bacillus subtilis)]	87	69	288
3081	2	126	308	911467399	[THF dehydrogenase (Bacillus subtilis)]	87	72	183
3535	1	3	401	911105554	[farnesyltransferase (Bacillus subtilis)]	87	80	399
4228	1	547	275	91103765	[HtrU protein, urocanase (Bacillus subtilis)]	87	73	273
4	8	10427	8736	91103769	[HtrU protein, urocanase (Bacillus subtilis)]	86	72	1692
22	6	4190	3738	911410515	[urease beta subunit (Staphylococcus xylosus)]	86	73	453
54	2	2480	1572	911203287	[UDP-glucose pyrophosphorylase (Bacillus subtilis)]	86	70	909
124	3	2336	1713	911558887	[uracil phosphoribosyltransferase (Bacillus subtilis - Bacillus subtilis)]	86	74	624
148	3	1349	1	911467458	[cell division protein (Bacillus subtilis)]	86	75	2100
149	4	13638	2859	911467460	[Unknown (Bacillus subtilis)]	86	73	222
152	3	1340	2086	911377835	[pyruvate decarboxylase E1-alpha subunit (Bacillus subtilis)]	86	75	747
164	18	117347	119467	9111184660	[polynucleotide phosphorylase (Bacillus subtilis)]	86	72	2121
180	2	1554	1159	911143467	[ribosomal protein S4 (Bacillus subtilis)]	86	80	606
205	3	2966	1	91142464	[ribosomal protein L17 (Bacillus subtilis)]	86	77	375
205	26	13364	12390	91140107	[ribosomal protein L22 (Bacillus stearothermophilus) irs[10612]9110612]	86	75	375
246	7	3463	3140	911467375	[ribosomal protein S6 (Bacillus subtilis)]	86	70	324
299	3	1196	1340	91139556	[ispQG gene product (Bacillus megaterium)]	86	70	345
299	7	3884	4345	911467440	[phosphoribosylpyrophosphate synthetase (Bacillus subtilis) synthetase (AA 1-377) (Bacillus subtilis)]	86	78	462
304	5	2170	2323	911665983	[putative ATP binding subunit (Bacillus subtilis)]	86	65	354
310	2	1487	1678	9111177684	[chorismate mutase (Staphylococcus xylosus)]	86	71	192
337	5	2086	3405	911467434	[isocitrate dehydrogenase (Bacillus subtilis)]	86	78	1320
339	2	1489	1109	911118003	[dihydroxyacetone kinase (Staphylococcus haemolyticus)]	86	77	381
358	2	2124	3440	911116219	[28.2% of identity to the Escherichia coli GTP-binding protein Era; putative (Bacillus subtilis)]	86	73	1317

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
404	12	1015	2058	0 1 1 303817	[YglA (Bacillus subtilis)]	66	78	1044
581	2	661	452	0 1 1 00056	[lipop gene product (Bacillus subtilis)]	86	71	210
612	2	138	1075	0 1 1 176399	[Spf (Staphylococcus epidermidis)]	86	72	738
770	1	622	247	0 1 1 43326	[lipop protein (putative) (Bacillus subtilis)]	86	69	276
855	1	1777	690	0 1 1 146247	[leucyltRNA synthetase (Bacillus subtilis)]	86	74	881
868	12	963	1133	0 1 1 002911	[transmembrane protein (Saccharomyces cerevisiae)]	86	69	171
904	1	1	622	0 1 1 303912	[YdhW (Bacillus subtilis)]	86	72	162
989	1	35	433	0 1 1 303993	[YohI (Bacillus subtilis)]	86	76	399
1212	1	296	150	0 1 1 4014	[ipa-90d gene product (Bacillus subtilis)]	86	70	147
1323	1	2	148	0 1 1 00441	- [pyruvate dehydrogenase (lipoprotein) (Bacillus stearothermophilus)] - [lipoyl-β-aldehyde-β-pyruvate dehydrogenase (lipoprotein) (Bacillus stearothermophilus)]	86	75	147
3085	2	540	310	0 1 1 354211	[PET112-like protein (Bacillus subtilis)]	86	86	231
1047	1	1	22R	0 1 1 296464	[ATPase (Lactococcus lactis)]	86	63	22R
4487	1	476	240	0 1 1 022726	[unknown (Staphylococcus haemolyticus)]	86	73	237
4583	1	372	187	0 1 1 022225	[unknown (Staphylococcus haemolyticus)]	86	79	186
25	5	4277	5039	0 1 1 502421	[3-ketoacyl-acyl carrier protein reductase (Bacillus subtilis)]	85	64	753
56	121	10637	29395	0 1 1 408507	[pyrimidine nucleotide transport protein (Bacillus subtilis)]	85	69	12,313
68	2	332	1192	0 1 1 467316	[unknown (Bacillus subtilis)]	85	74	861
73	2	860	1707	0 1 1 29392	[glycerol kinase (GK) (EC 2.7.1.30) (Bacillus subtilis) 0 1 1 5868 (B45568)]	85	72	828
106	4	1505	3490	0 1 1 33766	[thierry (EC 6.1.1.1) (Bacillus subtilis)]	85	74	1986
124	2	1133	2202	0 1 1 31924	[glyceraldehyde-3-phosphate dehydrogenase (Clostridium pasteurianum)]	85	75	1050
129	4	6466	5257	0 1 1 064807	[arginine amidotransferase (Bacillus subtilis)]	85	73	1215
138	6	3415	5673	0 1 1 072119	[glcB gene product (Staphylococcus carnosus)]	85	74	2199
189	1	2	169	0 1 1 467385	[unknown (Bacillus subtilis)]	85	65	168

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start ID	Stop ID	match accession	match gene name	% sim	% ident	length (nt)
205	15	8624	8106	gi 1046981	ribosomal protein S5 [Bacillus subtilis]	85	75	519
205	120	110328	110596	gi 103819 gb SBS	ribosomal protein L24 - Bacillus stearothermophilus	85	72	333
220	6	6490	6101	gi 148980	leach gene product [Bacillus subtilis]	85	66	390
231	4	4877	3159	gi 1100520	huts [Bacillus subtilis]	85	70	1719
243	9	8013	8763	gi 1114011	lpa-87r gene product [Bacillus subtilis]	85	72	771
249	2	5894	3106	gi 1405454	leconitase [Bacillus subtilis]	85	73	2709
302	1	140	475	gi 101773	homolog of E. coli ribosomal protein L21 [Bacillus subtilis] ir S18439 S18439 Ribosomal protein L21 - Bacillus subtilis p P26908 R121 BACSU SOS RIBOSOMAL PROTEIN L21 (BL20).	85	72	336
333	1	5445	2968	gi 142260	C1PC adenosine triphosphatase [Bacillus subtilis]	85	69	2478
364	6	6012	8196	gi 171784	C1P-like ATP-dependent protease binding subunit [Bos taurus]	85	68	2315
448	2	1392	1339	gi 105334	acetate kinase [Bacillus subtilis]	85	68	654
747	1	1251	653	gi 1173157	[orf-X: hypothetical protein: Method: conceptual translation supplied by author [Bacillus subtilis]]	85	73	399
886	2	159	467	gi 151768	[human peroxease (Yersinia enterocolitica)]	85	55	109
1089	1	1206	606	gi 184154 gb 71	[signal recognition particle S4K chain homolog Ffh - Bacillus subtilis]	85	71	603
1163	1	816	409	gi 104155	[diaminopimelate decarboxylase (Bacillus mathanolicus) sp P41023 DODA_BACM7 DIAMINOPIMELATE DECARBOXYLASE (EC 4.1.1.20) DAP DECARBOXYLASE].	85	62	408
1924	1	487	251	gi 1215098	[excisionase (Bacteriophage 154e)]	85	73	237
2932	1	776	390	gi 1041099	[Pyruvate kinase [Bacillus licheniformis]]	85	71	387
3030	1	3	275	gi 142370	[pyruvate formate-lyase (AA 1-760) (Escherichia coli) [gi 501781 S01788 FORMATE-C-ACETYLTRANSFERASE (EC 2.3.1.54) - <i>Escherichia coli</i>]]	85	74	273
3111	1	595	299	gi 163566	[limb deformity protein (Gallus gallus)]	85	85	297
3778	1	630	316	gi 1391840	[beta-subunit of HDT (Psuedomonas fragi)]	85	67	315
3835	1	1	187	gi 1204472	[type I restriction enzyme ECOR124/3 I M protein [Haemophilus influenzae]]	85	56	387
4042	1	3	386	gi 111178	[formate acetyltransferase (Escherichia coli) [gi 526991 S014997 FORMATE ACETYLTRANSFERASE (EC 2.3.1.54) - <i>Escherichia coli</i>]]	85	70	384
4053	1	35	340	gi 1204472	[formate C-acetyltransferase (EC 2.3.1.54) - <i>Escherichia coli</i> [gi 526991 S014997 FORMATE C-ACETYLTRANSFERASE (EC 2.3.1.54) - <i>Escherichia coli</i>]]	85	56	306
4108	1	2	181	gi 1072418	[glcA gene product [Staphylococcus carnosus]]	85	61	180
4300	1	575	330	gi 151932	[fructose enzyme II (Rhodobacter capsulatus)]	85	59	246

TABLE 2

EP 0 786 519 A2

S. aurous - putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start	Stop	match	match gene name	% sim	% ident	length
10	10	[nt]	[nt]	accession	r			[nt]
4392	1	627	355	[gi 1032725]	Unknown [Staphylococcus haemolyticus]	85	74	273
4408	1	2	235	[gi R71784]	[Ctn-like ATP-dependent protease binding subunit [Bos taurus]	65	62	234
4430	1	538	291	[gi 1009366]	Respiratory nitrate reductase [Bacillus subtilis]	85	68	288
4555	1	2	253	[gi 450688]	[auth gene of Ecoprt gene product [Escherichia coli] pir S01837 S01837 hadR	85	52	252
					[protein - Escherichia coli pir S01629 S01629 hypothetical protein A - Escherichia coli ISUS 40-320]			
4611	1	491	242	[gi 1236635]	[dihydroxy-acid dehydrogenase [Bacillus subtilis]	85	65	240
4	10	10061	10591	[gi 46982]	[tadB gene product [Staphylococcus epidermidis]	84	68	531
13	1	2	1348	[gi 142450]	[shcC protein [Bacillus subtilis]	84	56	177
16	1	4	1803	[gi 1227198]	[DNA repair protein [Bacillus subtilis radiodurans]	84	67	2850
22	1	1535	1128	[gi 1511069]	[ureF [Staphylococcus xylosus]	84	73	408
23	1	7	5055	[gi 1603220]	[YerOBp [Saccharomyces cerevisiae]	84	63	252
53	111	11597	11145	[gi 11301948]	[YQW [Bacillus subtilis]	84	68	453
53	112	14059	112770	[gi 142613]	[branched chain alpha-keto acid dehydrogenase E2 [Bacillus subtilis] gtr1303946 BtrB8 [Bacillus subtilis]	84	71	1270
70	1	1	1322	[gi 46647]	[ORF (repE) [Staphylococcus aureus]	84	68	351
73	1	4	2512	[gi 143993]	[glycerol-3'-phosphate dehydrogenase (gppD) [Bacillus subtilis]	84	74	1800
9R	1	7	4324	[gi 6096]	[methylonyl-rRNA synthetase [Bacillus subtilis]	84	66	1773
100	1	9	9501	[gi 8860]	[ORF1 [Staphylococcus aureus]	84	78	822
117	1	3	1934	[gi 1237019]	[srB [Bacillus subtilis]	84	68	1275
148	1	6	4720	[gi 5670]	[cysteine synthetase A [Bacillus subtilis]	84	69	951
152	4	2064	2456	[gi 143377]	[pyruvate decarboxylase (E-1) alpha subunit [Bacillus subtilis] pir B57101B83 hydrolase dehydrogenase (lipoprotein) (EC 1.2.4.11) pha	84	70	393
169	1	7	3634	[gi 1001342]	[hypothetical protein [Synechocystis sp. PCC 6803]	84	66	228
171	1	4	2992	[gi 1517475]	[D-alino acid transaminase [Staphylococcus haemolyticus]	84	71	336
186	1	6	6941	[gi 6216]	[Unknown [Bacillus subtilis]	84	70	726
205	1	9	6261	[gi 5692]	[ORF for adenylyl kinase [Bacillus subtilis]	84	71	570
224	1	2	915	[gi 1391]	[beta-fructofuranosidase [Staphylococcus xylosus]	84	70	477

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	orf	Start	Stop	match	match gene name	% sim	% ident	length
	ID	(ml)	(ml)	acquisition		(nt)	(nt)	(nt)
251	1	92	348	9111303790	[Ygi] (Bacillus subtilis)	84	65	297
282	3	1526	2036	911141040	[glutamate-1-semialdehyde 2,1-minotransferase (Bacillus subtilis)]	84	75	1311
					[pir102728] (Bacillus subtilis)			
					[D2728 glutamate-1-semialdehyde 2,1-minotransferase (EC 4.3.3.8)]			
307	5	3138	2959	9111070014	[protein-dependent (Bacillus subtilis)]	84	62	180
320	4	2343	4229	911143390	[carboxyl phosphate synthetase (Bacillus subtilis)]	84	70	1887
					[unknown (Staphylococcus haemolyticus)]	84	70	294
413	2	2201	1341	9111256146	[YbgQ (Bacillus subtilis)]	84	65	861
439	1	3	392	9111016173	[osmotically inducible protein (Mycoplasma genitalium)]	84	53	390
461	3	1362	2270	911140211	[threonine synthase (thrC) (IAA 1-152) (Bacillus subtilis) trc (Bacillus subtilis)]	84	69	909
					[threonine synthase (EC 4.2.99.2) - Bacillus subtilis]	84	46	297
487	1	3	299	9111144521	[lineatin-like protein alpha Int1p (Candida albicans)]	84	69	282
491	2	624	905	9111085641RBS	[ribosomal protein S9 - Bacillus stearothermophilus]	84	77	196
491	3	816	1013	9111085641RBS	[ribosomal protein S9 - Bacillus stearothermophilus]	84	74	339
544	1	3	341	9111411231	[inac1 peptidase (Bacillus caldolyticus)]	84	68	954
728	2	2701	1748	9111912445	[DNA polymerase (Bacillus caldotenax)]	84	57	151
769	1	3	257	9111510953	[cobalamin biosynthesis protein N (Methanococcus jannaschii)]	84	70	255
954	1	308	156	9111405454	[aconitase (Bacillus subtilis)]	84	56	255
957	1	3	395	911143402	[recombination protein (trg start codon) (Bacillus subtilis) gtr (Bacillus subtilis)]	84	68	393
975	1	3	452	911185934	[CluB (Synechococcus sp.)]	84	73	321
1585	1	3	257	9111510140	[lipoendopeptidase F (Lactococcus lactis)]	84	65	450
2954	1	3	323	9111603769	[HtrU protein, urocate (Bacillus subtilis)]	84	71	309
2996	1	650	348	91118178	[formate acetyltransferase (Chlamydomonas reinhardtii) trf (S24971) (S221997) (Chlamydomonas reinhardtii)]	84	77	246
					[formate C-acetyltransferase (EC 2.3.1.54) - Chlamydomonas reinhardtii]	84	72	363
3766	1	737	4375	911171205	[167 kDa Myo-inositol-crostatecile streptococcal antigen (Streptococcus pyogenes)]	84	54	168
4022	1	2	169	9111146206	[glutamate dehydrogenase (Bacillus subtilis)]	84	71	309
4058	1	620	312	911151932	[fructose enzyme II (Rhodobacter capsulatus)]	84	77	246
4108	2	196	351	9111072418	[glcK gene product (Staphylococcus carnosus)]	84	77	246

TABLE 2

5
10
15
20
25
30
35
40
45
50
55

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4183	1	1	308	gi 603769	[lucu] protein, urocanase [Bacillus subtilis]	84	72	306
4726	1	55	234	gi 146208	[glutamate synthase large subunit (EC 2.6.1.51) (Ecobrillia coll) (EC 1.4.1.13) large chain - Escherichia coli]	84	73	180
22	4	2043	1576	gi 139297	[urease accessory protein [Bacillus sp.]	83	64	468
53	113	14722	13745	gi 142612	[branched chain alpha-keto acid dehydrogenase E1-beta [Bacillus subtilis]]	83	68	978
57	116	13357	12872	gi 143132	[lactate dehydrogenase (EC 1.1.1.27) [Bacillus caldolyticus] (EC 1.1.1.27) - Bacillus caldolyticus]	83	66	486
66	3	3119	2274	gi 1303894	[YdhM [Bacillus subtilis]]	83	63	846
66	5	6118	4643	gi 11212730	[YdhK [Bacillus subtilis]]	83	68	1476
70	3	1864	1523	gi 144095	[replication initiator protein [Listeria monocytogenes]]	83	73	342
90	1	377	1429	gi 1555571	[alcohol dehydrogenase I (adhB) (EC 1.1.1.1) (Zymomonas mobilis) (EC 1.1.1.1) I - Zymomonas mobilis]	83	70	1053
95	2	708	2162	gi 1506381	[l-phospho-beta-glucosidase [Bacillus subtilis]]	83	70	1455
127	1	68	64	gi 1467391	[initiation protein of replication [Bacillus subtilis]]	83	77	627
140	4	3209	2742	gi 163107	[kdpB [Escherichia coli]]	83	65	468
142	3	3466	2869	gi 11212776	[lumazine synthase (b-subunit) [Bacillus amyloliquaticeps]]	83	69	480
161	112	5749	6696	gi 1903307	[orfF5 [Bacillus subtilis]]	83	64	948
164	9	9680	11070	gi 149316	[orfF2 gene product [Bacillus subtilis]]	83	66	1191
164	114	14148	14546	gi 11580902	[orfF6 gene product [Bacillus subtilis]]	83	60	399
170	1	3144	2467	gi 1520844	[orf4 [Bacillus subtilis]]	83	64	678
186	1	2029	1370	gi 289284	[cytosanyl-tRNA synthetase [Bacillus subtilis]]	83	72	660
205	14	7622	7607	gi 1216137	[orf for L10 ribosomal protein [Bacillus subtilis]]	83	74	216
237	6	3683	4540	gi 1510486	[isidazoleglycerol-phosphate synthase (cyclase) [Methanococcus jannaschii]]	83	60	858
301	1	985	638	gi 467419	[unknown [Bacillus subtilis]]	83	65	348
302	4	1421	2743	gi 1508979	[GTP-binding protein [Bacillus subtilis]]	83	68	1323
321	4	3933	3571	gi 139844	[fumurate reductase (EC 1.46.2) [Bacillus subtilis]]	83	68	363
367	1	2	352	gi 1039479	[orfU [Lactococcus lactis]]	83	54	351

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
313	1	355	179	[gi 142455]	[alanine dehydrogenase (EC 1.4.1.1) (Bacillus stearothermophilus)]	83	66	177
387	1	3	662	[gi 066281]	[DNA polymerase I (Bacillus stearothermophilus)]	83	70	660
527	2	916	1566	[gi 139659]	[protease (Staphylococcus epidermidis)]	83	67	651
533	1	2	859	[gi 1520753]	[DNA topoisomerase I (Bacillus subtilis)]	83	72	858
774	2	200	361	[gi 1522645]	[N- Jammachil predicted coding region (Bacillus subtilis)]	83	58	162
897	1	120	296	[gi 1064807]	[Ornithine Aminotransferase (Bacillus subtilis)]	83	76	177
1213	1	3	491	[gi 1492888]	[LysA (Bacillus subtilis)]	83	67	489
2529	1	296	150	[gi 1437786]	[tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Bacillus subtilis)]	83	69	147
2873	1	649	326	[gi 1105687]	[Prot (Bacillus subtilis)]	83	58	324
3009	1	728	366	[gi 182532]	[ORF_0294 (Escherichia coli)]	83	65	363
3015	2	45	305	[gi 1950062]	[hypothetical yeast protein 1 (Hypoplasma capricolum pir[S48578]S18578] (fragment)]	83	59	261
3006	1	67	309	[gi 1353197]	[Chlorodin reductase (Escherichia acidaminophilum)]	83	61	243
4458	1	540	1271	[gi 1397526]	[clumping factor (Staphylococcus aureus)]	83	78	270
4570	1	444	223	[gi 1022726]	[Unknown (Staphylococcus haemolyticus)]	83	74	222
4854	1	97	261	[gi 1072419]	[glcB gene product (Staphylococcus carnosus)]	83	79	165
16	2	295	1191	[gi 153854]	[lvs002 protein (Streptococcus pneumoniae)]	82	67	897
16	3	1393	1798	[gi 153854]	[lvs002 protein (Streptococcus pneumoniae)]	82	70	606
18	12	9644	8724	[gi 1204400]	[N-acetylneuraminate lyase (Neisseria influenzae)]	82	58	921
42	4	988	2019	[gi 841192]	[catalase (Bacteroides fragilis)]	82	70	1012
51	6	2520	3489	[gi 143607]	[sporulation protein (Bacillus subtilis)]	82	69	900
56	11	12270	13925	[gi 139431]	[oligo-1,6-glucosidase (Bacillus cereus)]	82	60	1056
56	15	17673	18014	[gi 467410]	[Unknown (Bacillus subtilis)]	82	66	342
61	2	881	3333	[gi 131448]	[Transfer RNA-Lau synthetase (Bacillus subtilis)]	82	70	2433

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	orf	Start	Stop	match	match	patch	gene name	% sim	% ident	length	
ID	ID	Int1	Int1	accession	accession			Int1	Int1	Int1	
62	7	9162	11318	gi 48240	[gi 48240]	elongation factor G (AA 1-691) [Thermus aquaticus thermophilus]		82	64	2157	
						[gi 52281 EFM03 translation elongation factor G - Thermus aquaticus					
						orf1535 EFM03 elongation factor G (EF-G).					
65	1	5470	3260	gi 143369	[gi 143369]	phosphoribosylformyl glycaminidine synthetase II (PUR-Q)	[Bacillus subtilis]	82	66	2211	
102	6	1662	5380	gi 1256635	[gi 1256635]	1,4-dihydroxy-acid dehydrogenase [Bacillus subtilis]		82	65	1719	
117	4	3202	1493	gi 471541 AT471	[gi 471541 AT471] orf1 5' of Rfb - Bacillus subtilis			82	53	252	
128	6	4377	5933	gi 460258	[gi 460258]	phosphoglycerate mutase [Bacillus subtilis]		82	66	1557	
129	2	1229	2182	gi 403373	[gi 403373]	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis]		82	62	954	
						[gi 517251 S7251 glycerophosphoryl diester phosphodiesterase - <i>actinillus</i>					
						subtilis					
170	1	2	1441	gi 1377831	[gi 1377831]	unknown [Bacillus subtilis]		82	67	1440	
177	1	3	1094	gi 1467386	[gi 1467386]	thiophen and furan oxidation [Bacillus subtilis]		82	65	1092	
184	1	4	3572	4039	gi 153566	[gi 153566]	ORF 119K protein [Enterococcus faecalis]		82	59	468
189	1	R	4455	4225	gi 1001978	[gi 1001978]	Cp01 protein [Listeria monocytogenes]		82	73	231
206	19	21366	20707	gi 473916	[gi 473916]	lipopeptide antibiotics iturin A [Bacillus subtilis] sp19314 LP14_BACSU		82	50	660	
						LIPOPERMIDE ANTIBIOTICS ITURIN A AND SURFACTIN ISOENYMES IS PROTEIN					
221	2	805	1722	gi 1517205	[gi 1517205]	67 kDa Mysin-crossreactive streptococcal antigen [Streptococcus yoghensis]		82	63	918	
233	4	3866	3651	gi 439619	[gi 439619]	[Salmonella typhimurium] IS200 insertion sequence from SAR17, attab-1, gene product [Salmonella typhimurium]		82	69	216	
260	3	5207	4296	gi 1161181	[gi 1161181]	[Staphylococcus epidermidis]		82	61	912	
315	1	4864	2853	gi 143397	[gi 143397]	[quino] oxidase [Bacillus subtilis]		82	67	2010	
321	110	8520	7945	gi 1142981	[gi 1142981]	ORFs: This ORF includes a region (aa23-101) containing a potential rom-sulphur centre homologous to a region of Rhodospirillum rubrum nd-Chromatium vinosum; putative [Bacillus stearothermophilus] pfr P02399 P02399 hypothetical protein 5 [gi da3] - region1 -		82	62	576	
	3	1055	1342	gi 436574	[gi 436574]	ribosomal protein L1 [Bacillus subtilis]		82	71	288	
	370	2	262	618	[gi 1103793]	[YqeL [Bacillus subtilis]]		82	59	357	
404	4	3053	4024	gi 1103021	[gi 1103021]	[YqeE [Bacillus subtilis]]		82	68	972	
405	4	4400	3073	gi 1103913	[gi 1103913]	[YohX [Bacillus subtilis]]		82	67	1160	
436	3	4096	2864	gi 149521	[gi 149521]	tryptophan synthase beta subunit [Lactococcus lactis] pfr S35129 S35129		82	67	1233	
						tryptophan synthase alpha chain (EC 4.2.1.20) - actococcus lactis subsp.					
						lactis					

TABLE 2

S. aurous - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
441	1	4	394	gi 11227	glyceraldehyde-3-phosphate dehydrogenase (Bacillus stearothermophilus)	82	67	822
444	12	110415	111227	gi 11201354	spore germination and vegetative growth protein (Bacillus stearothermophilus)	82	67	813
446	1	3	191	gi 1143387	aspartate transcarbamoylase (Bacillus subtilis)	82	66	189
462	3	1007	1210	gi 1142521	deoxyribodipyrimidine photolyase [Bacillus subtilis] pir A37192 A37192_UVRE_BACSU EXCINUCLEASE ABC SUBUNIT C.	82	64	204
517	1	1560	1784	gi 11653767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (Bacillus subtilis)	82	61	777
660	2	407	700	gi 11426472	lsecA gene product (Staphylococcus carnosus)	82	69	294
724	2	565	186	gi 1143373	phosphoribosyl aminimidazole carboxy-lyase (IPUR-R17) Bacillus subtilis	82	68	180
763	1	422	213	gi 11467458	cell division protein (Bacillus subtilis)	82	35	210
818	1	564	1283	gi 11064787	function unknown (Bacillus subtilis)	82	69	282
858	1	175	1176	gi 1143043	lucoporphyrinogen decarboxylase (Bacillus subtilis) pir A704 B7045	82	71	1002
895	1	3	599	gi 1025507	lucoporphyrinogen decarboxylase (EC 4.1.1.37) - acillus subtilis	82	72	597
919	1	10	1399	gi 1143795	ATP binding protein [Borella burgdorferi]	82	60	396
961	1	1	306	gi 11576847	ltranser RNA-Tyr synthetase (Bacillus subtilis)	82	69	306
1192	1	107	155	gi 1145974	[NH3]-dependent NAD synthetase [Escherichia coli]	82	71	153
1317	1	49	1375	gi 1107908	leucine tRNA synthetase (Staphylococcus xylosus)	82	72	327
1341	1	1	150	gi 1139962	ribosomal protein L5 (AA 1-66) [Bacillus stearothermophilus] ir S0537 B5835	82	68	150
2990	2	567	349	gi 1134655	ribosomal protein L5 (Bacillus stearothermophilus) sp P42009 ATPE_BACST	82	47	219
3024	1	45	224	gi 1167402	[ATP synthase epsilon chain (EC 3.6.1.34)]	82	64	160
3045	1	276	139	gi 1167335	lunknown (Bacillus subtilis)	82	60	138
3045	2	558	400	gi 1167335	lribosomal protein L9 (Bacillus subtilis)	82	82	159
3091	1	474	238	gi 1199335	lsecA protein (Staphylococcus carnosus)	82	78	237
3107	1	416	210	gi 1546318	lory Y 1- of comK [Bacillus subtilis], E26, Peptide Partial, 140 aa] pir S1362 S1362 hypothetical protein Y - Bacillus subtilis sp P0396 YHD_BACSU HYPOTHETICAL PROTEIN IN COMK 3 REGION (ORF7).	82	64	207

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	Length (nt)
4332	1	2	319	gi 142086	nitrate reductase alpha subunit (Escherichia coli) DIP0152 NARD_ECOLI	82	75	318
23	3	3275	2574	gi 11129573	RESPIRATORY NITRATE REDUCTASE 1 ALPHA CHAIN (EC 7.99.4) (SUB 2-1247)	81	64	702
42	1	638	321	gi 146778	Lysine specific permease (Escherichia coli)	81	59	318
48	5	4051	4350	gi 105937	N- genitilium predicted coding region NC216 [Mycoplasma genitalium]	81	62	300
51	4	1578	2579	ipk15_669 S166	iclAC protein - Bacillus subtilis	81	55	1002
53	2	354	1494	gi 1130961	[tpqJ] Bacillus subtilis	81	67	1131
53	8	9419	7971	gi 146930	6'-phosphogluconate dehydrogenase (Escherichia coli)	81	66	1449
54	9	10757	10119	gi 141016	permease [Bacillus subtilis]	81	65	639
54	10	13360	11786	gi 143015	gluconate kinase [Bacillus subtilis]	81	66	1575
57	117	13983	13366	ipk15_805 A258	L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus subtilis	81	74	618
81	2	2708	2217	gi 11222302	NIKU-related protein [Hemophilus influenzae]	81	54	492
86	1	745	374	gi 141017	[tpa-910] gene product [Bacillus subtilis]	81	70	372
103	6	6328	4861	gi 1971342	NITRATE REDUCTASE BETA CHAIN (EC 1.7.99.41)	81	64	1576
120	15	110845	12238	gi 1524392	[gb3A] [Bacillus subtilis]	81	67	1494
128	5	13676	4413	gi 143319	triose phosphate isomerase [Bacillus megaterium]	81	64	738
131	9	10308	9280	gi 1293163	laminine dehydrogenase [Bacillus subtilis]	81	68	1029
143	6	6088	5671	gi 1432619	Isolamella typhimurium IS200 insertion sequence from SARA17, attI.1, gene product [Salmonella typhimurium]	81	61	618
169	1	43	825	gi 89795	[30S ribosomal protein (pediococcus acidilactici)] sp P49668 RS2_PEDIC_30S	81	65	783
230	1	450	226	gi 1125086	RIBOSOMAL PROTEIN S2.	81	54	225
233	5	2000	2677	gi 1467404	short region of weak similarity to tyrosine-protein kinase receptors in <i>A. thaliana</i> (fragment)	81	63	678
241	2	1081	2149	gi 116310	succinate-CoA ligase (GDP-forming) [Arabidopsis thaliana] (EC 6.2.1.4) pha chain - Arabidopsis thaliana (fragment)	81	69	933
256	1	1	981	gi 15941115094	spolE protein - Bacillus subtilis	81	65	981
259	3	3752	2691	ipk15_8057 RF2_2	[PROBABLE PEPTIDE CHAIN RELEASE FACTOR 2 (RF-2) (FRAGMENT)]	81	65	1052

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
275	2	1728	3581	[gi 776480	[L-glutamine-D-fructose-6-phosphate amidotransferase (Bacillus subtilis)]	81	68	1854
285	1	1466	1735	[gi 1204844	[H. influenzae predicted coding region Hf0594 (Haemophilus influenzae)]	81	63	732
296	1	99	1406	[gi 467228	[adenylosuccinate synthetase (Bacillus subtilis)]	81	67	1308
302	9	5590	5889	[gi 147485	[queA (Escherichia coli)]	81	64	300
317	2	1137	1376	[gi 154961	[resolvase (transposon Tn917)]	81	51	240
343	2	1034	1342	[gi 405955	[yed (Escherichia coli)]	81	60	309
360	2	1404	2471	[gi 1201570	[aspartyl-tRNA synthetase (Haemophilus influenzae)]	81	67	1068
364	5	6251	5706	[gi 1204652	[methylated-DNA-protein-cysteine methyltransferase (Haemophilus influenzae)]	81	63	546
372	2	1707	1135	[gi 467416	[unknown (Bacillus subtilis)]	81	65	573
392	1	43	603	[pir 50411 S094	[sporite protein - Bacillus subtilis]	81	65	561
404	9	5252	6154	[gi 606745	[Rex (Bacillus subtilis)]	81	65	903
426	2	1727	1119	[gi 39453	[Manganese superoxide dismutase (Bacillus caldotenax) tr S22053 S22053]	81	66	609
440	7	1653	5849	[gi 1370	[hypothetical protein II (lmp13) region] - <i>Salmonella typhimurium</i>	81	57	217
625	3	1105	2070	[gi 1262360	[protein kinase PknB (Hydrobacterium leprae)]	81	56	966
754	2	504	1064	[gi 1303902	[Qnu (Bacillus subtilis)]	81	71	561
842	1	86	430	[gi 405446	[transktolase (Bacillus subtilis)]	81	68	345
933	1	798	400	[gi 205429	[dipeptide transport ATP-binding protein (Haemophilus influenzae)]	81	57	399
961	2	252	401	[gi 487686	[synergohemagglutinotropic toxin (Staphylococcus intermedius) tr S494 S4944]	81	72	150
					[synergohemagglutinotropic toxin - <i>Staphylococcus nigericus</i>]			
1015	1	1	189	[gi 1046138	[H. genitalium predicted coding region Hg423 (Mycoplasma genitalium)]	81	43	189
1280	1	670	449	[gi 529164	[helicase (Autographa californica nuclear polyhedrosis virus) sp P2407 V43_NFVAC HELICASE]	81	43	222
3371	1	68	241	[gi 322245	[levulonate myrophosphate decarboxylase (Rattus norvegicus)]	81	62	174
3315	1	475	239	[gi 527137	[ORF_1388 (Escherichia coli)]	81	58	237
3908	1	2	325	[gi 439619	[Salmonella typhimurium IS2200 insertion sequence from SAR17, attab1, gene product (Salmonella typhimurium)]	81	68	324
3940	1	3	401	[gi 216664	[ATPase (Lactococcus lactis)]	81	69	399

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match expression	match gene name	% sim	% ident	length (nt)
3934	1	1	318	[91]1224069	[imidazolecatathialase]	81	68	316
4049	1	337	170	[91]1603768	[ut1 protein, imidazolone-5-propionate hydrolase <i>Bacillus subtilis</i> 91]603768 Nut1 protein, imidazolone-5-propionate hydrolase <i>Bacillus subtilis</i>]	81	68	168
4269	1	324	4103373		[glycerophosphoryl diester phosphodiesterase (Bacillus subtilis) - acillus subtilis sp1537251] [S1537251 glycerophosphoryl diester phosphodiesterase - acillus subtilis]	81	58	324
4371	1	627	122	[91]216677	[indolepyruvate decarboxylase (Enterobacter cloacae) pIR156013] [S156013 indolepyruvate decarboxylase (EC 4.1.1.-) - <i>Enterobacter cloacae</i>]	81	72	306
4387	1	19	228	[91]460669	[TVG (Thermococcus vulgaris)]	81	59	210
4191	1	581	306	[91]1524193	[unknown (<i>Mycobacterium tuberculosis</i>)]	81	67	276
4425	1	3	341	[91]113015	[gluconate kinase (<i>Bacillus subtilis</i>)]	81	66	339
9	1	1593	647	[91]1064786	[function unknown (<i>Bacillus subtilis</i>)]	80	62	747
17	1	544	311	[91]559164	[helicase (<i>Autographa californica</i> nuclear polyhedrosis virus) sp1P24307 V143_NPVAC_Helicase]	80	40	234
45	2	1159	2448	[91]110964	[ProV (<i>Bacillus subtilis</i>)]	80	63	1250
45	5	4032	4733	[91]1109887	[ProZ (<i>Bacillus subtilis</i>)]	80	55	702
54	8	110266	9502	[91]1563952	[gluconate permease (<i>Bacillus licheniformis</i>)]	80	62	765
62	12	8832	7545	[91]844655	[Na/H antiporter system (<i>Bacillus alcalophilus</i>)]	80	62	1308
62	14	8087	6683	[91]1559713	[ORF (<i>luteo splenii</i>)]	80	68	597
67	16	13781	14122	[91]105002	[ORF_E356 (<i>Escherichia coli</i>)]	80	65	342
70	13	11495	10296	[91]1103995	[YqKN (<i>Bacillus subtilis</i>)]	80	64	1200
98	15	6326	7130	[91]147428	[unknown (<i>Bacillus subtilis</i>)]	80	68	795
98	10	7294	7833	[91]1467430	[unknown (<i>Bacillus subtilis</i>)]	80	64	540
98	11	7820	8737	[91]1467431	[high level kanamycin resistance (<i>Bacillus subtilis</i>)]	80	61	918
109	16	14154	14813	[91]150875	[lpp57d gene product (<i>Bacillus subtilis</i>)]	80	63	660
112	15	14294	146636	[91]1072261	[pyruvate-formate-lyase (<i>Clostridium pasteurianum</i>)]	80	65	2343
139	1	1446	1726	[91]1506639	[CapC (<i>Staphylococcus aureus</i>)]	80	58	723
139	2	2179	1448	[91]1506638	[CapB (<i>Staphylococcus aureus</i>)]	80	59	722
174	4	3271	2870	[91]1146242	[aspartate 1-decarboxylase (<i>Bacillus subtilis</i>)]	80	61	402

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (Int)	Stop (Int)	match accession	match gene name	% sim	% ident	length (nt)
186	1	5368	3875	[gi 1289282]	[glutamyl-tRNA synthetase (Bacillus subtilis)]	80	70	741
205	130	15796	15140	[gi 10103]	[ribosomal protein L4 (Bacillus stearothermophilus)]	80	66	657
207	1	140	1315	[gi 160259]	[lenolase (Bacillus subtilis)]	80	61	1176
211	3	1078	1390	[gi 101031]	[ORF7 (Bacillus subtilis)]	80	61	513
235	2	1962	2255	[gi 143797]	[valyl-tRNA synthetase (Bacillus stearothermophilus) sp P1931 ISV_BACST (VALASE) (VALIS)]	80	55	294
239	1	1	1263	[gi 143000]	[proton glutamate symporter protein (Bacillus stearothermophilus) sp S26247 S26247 glutamate/aspartate transport protein - Bacillus stearothermophilus]	80	59	1263
272	5	2724	2461	[gi 109993]	[hypothetical protein (Bacillus subtilis)]	80	54	264
301	3	1445	1111	[gi 1467418]	[unknown (Bacillus subtilis)]	80	58	336
310	4	1567	14501	[gi 1177686]	[lucu gene product (Staphylococcus xylosus)]	60	67	1197
310	6	5258	7006	[gi 148053]	[acetyl-CoA synthetase (Bacillus subtilis)]	80	67	1749
310	7	7410	9113	[gi 1103865]	[farnyl-tetrahydrofolate synthetase (Streptococcus mutens)]	80	67	1704
325	3	1114	1369	[gi 10325]	[outer capsid protein (Rotavirus sp.)]	80	40	276
337	1	1268	616	[gi 157049]	[ORF_0470 (Escherichia coli)]	60	55	633
374	2	929	1228	[gi 1405448]	[lneF (Bacillus subtilis)]	80	70	300
375	5	3062	3331	[gi 1467448]	[unknown (Bacillus subtilis)]	80	68	270
388	1	267	587	[gi 1064791]	[function unknown (Bacillus subtilis)]	80	65	121
394	1	9	659	[gi 104976]	[matches P00017: ATP_GTP_A and P50001: EFACTOR_GTP; similar to elongation factor G, TetM/TetD tetracycline-resistance protein Escherichia coli]	80	65	651
456	1	625	1263	[gi 1146183]	[putative (Bacillus subtilis)]	80	65	619
475	1	1	654	[gi 1288269]	[beta-fructofuranosidase (Staphylococcus xylosus)]	80	66	654
544	2	1449	2240	[gi 529754]	[spec (Streptococcus pyogenes)]	80	50	792
622	4	1623	1871	[gi 1483945]	[unknown (Mycobacterium tuberculosis)]	80	65	249
719	1	1	1257	[gi 1064791]	[function unknown (Bacillus subtilis)]	80	68	1257
739	1	107	838	[gi 166693]	[putative ATP binding subunit (Bacillus subtilis)]	1	80	61
								732

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

contig	ORF ID	Start	Stop	match accession	match gene name	% sim	% ident	length (nt)
745	1 2	581	414	gi 1511600	[cysteine PQQ synthesis protein III (Methanococcus jannaschii)]	80	61	168
822	1 1	17	679	gi 410141	[orfX17 (Bacillus subtilis)]	80	68	661
827	1 2	991	836	gi 1205301	[leukotoxin secretion Ape-binding protein (Hemophilus influenzae)]	1	80	54
1044	1 1	3	149	gi 60632	[vp2 (Marburg virus)]	1	80	55
1220	1 2	571	413	ipr A61072 ESG	[leidlerin precursor - Staphylococcus gallinarum]	1	80	78
2519	1 1	75	275	gi 147556	[ipj (Escherichia coli)]	80	45	201
2947	1 1	503	279	gi 1184880	[polynucleotide phosphorylase (Bacillus subtilis)]	80	62	225
3120	1 1	2	226	gi 1517205	[67 kDa Hyaluron-crossreactive streptococcal antigen (Streptococcus yoghensis)]	1	80	65
3171	1 1	294	148	gi 151259	[HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas fluorescens) flrl44756 NA4756]	80	59	147
3560	1 2	285	434	gi 1217130	[hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.]	80	70	150
3655	1 1	47	346	gi 1415855	[photosystem I core protein B (Synechococcus vulgaris)]	1	80	56
3656	1 2	324	584	gi 151551	[deoxyribose aldolase (Hycoplasma hominis)]	1	80	50
3769	1 1	798	400	gi 1339950	[2-nitropropene dioxygenase (Methylospora saturni)]	1	80	54
3781	1 1	692	348	gi 166412	[large subunit of NADH-dependent glutamate synthase (Medicago sativa)]	1	80	62
3948	1 1	48	287	gi 1204696	[NADH-glutamate synthase (Medicago sativa)]	1	80	345
4030	1 1	571	287	gi 11009366	[fructose-6-phosphate 1,6-eliminase (Bacillus subtilis)]	80	69	240
4092	1 1	547	275	gi 11270207	[respiratory nitrate reductase (Bacillus subtilis)]	80	60	285
4103	1 1	680	342	gi 139956	[respiratory nitrate reductase (Bacillus subtilis)]	80	69	273
4231	1 1	692	348	gi 1289287	[respiratory nitrate reductase (Bacillus subtilis)]	80	65	249
4265	1 1	595	299	gi 603768	[HutI protein, imidazolone-5-propionate hydrolase (Bacillus subtilis) q1 603768 HutI protein, imidazolone-5-propionate hydrolase (Bacillus subtilis)]	80	63	297
4504	1 1	498	250	gi 1339950	[large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)]	80	68	249
2	6	5998	6798	gi 1535351	[Cody (Bacillus subtilis)]	1	79	63
4	7	8295	7057	gi 1603768	[HutI protein, imidazolone-5-propionate hydrolase (Bacillus subtilis) q1 603768 HutI protein, imidazolone-5-propionate hydrolase (Bacillus subtilis)]	79	64	1245
25	6	5273	5515	ipr A36728 A37	[acyl carrier protein - Rhizobium meliloti]	1	79	65

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (Int)	Stop (Int)	match accession	match gene name	% sim	% ident	length (Int)
59	2	1173	1424	gi 147923	[threonine dehydratase 2 (EC 4.2.1.16) [Escherichia coli]	79	75	252
60	1	1	204	gi 666115	[orf1 upstreem of glucose kinase [Staphylococcus xylosus] pir 552151 582351]	79	60	204
81	1	3002	1590	gi 466882	[hypothetical protein 1 - Staphylococcus xylosus	79	64	1413
85	7	7023	6505	gi 143364	[ppb1; B169_C2_169 [Mycobacterium leprae]	79	60	519
89	6	5660	4554	gi 144906	[phosphoribosyl aminimidazole carboxylase I (PUR-E) [Bacillus ubilliae]	79	35	1107
102	111	7489	8571	gi 143093	[product homologous to E. coli thiorodorn reductase; J. Biol. Chem. 1988] 263:9015-9019, and to P22 protein of alkyl hydroperoxide reductase from S. Typhimurium; J. Biol. Chem. 119901 265(1):0535-10340; pen reading frame A [Clostridium pasteurianum]	79	64	1033
102	14	11190	12563	gi 149728	[ketol-acid reductoisomerase [Bacillus subtilis] pir P37251 11VC. BACSU KETOL-ACID REDUCTOISOMERASE (EC 1.1.1.86) ACETOKYDROXY-ACID ISOMEROREDUCTASE] (ALPHA-KETO-BETA-HYDROXYLACIL EDUCTOISOMERASE).	79	65	1374
127	9	7792	9372	gi 458688	[PrfC/RP3 [Bacillus nodosus]	79	68	1561
139	1	250	1983	gi 506597	[CapA [Staphylococcus aureus]	79	55	558
144	2	1644	1156	gi 149296	[peptide methionine sulfoxide reductase [Streptococcus pneumoniae]	79	47	489
148	2	529	1098	gi 467157	[hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis] gi 467457 hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis]	79	59	570
150	1	965	591	gi 1755602	[unknown [Bacillus subtilis]	79	61	375
176	1	1039	507	gi 297874	[fructose-bisphosphate aldolase [Staphylococcus carnosus] pir A4994 A4994] (fructose-bisphosphate aldolase (EC 4.1.2.11)) - taphylococcus carnosus (strain M100)	79	65	453
186	7	7584	6874	gi 1311298	[ORF5: putative Ssa protein; similar to Ssa proteins from Haemophilus influenzae and Escherichia coli (Listeria monocytogenes)]	79	64	711
205	116	8497	8498	gi 1044980	[fructose-bisphosphate aldolase [Bacillus subtilis] (ribosomal protein L18 [Bacillus subtilis])	79	70	390
211	1	1	519	gi 1303994	[YQH [Bacillus subtilis]]	79	62	519
223	2	4183	2801	gi 488410	[alcohol dehydrogenase 2 [Entamoeba histolytica]	79	60	1363
243	6	8915	7896	gi 540883	[lpa-8id gene product [Bacillus subtilis]]	79	60	1020
279	4	3721	4329	gi 413930	[lpa-6d gene product [Bacillus subtilis]]	79	59	609
300	1	11	1393	gi 403372	[glycerol 3-phosphate permease [Bacillus subtilis]]	79	62	1383
307	3	2930	1935	gi 193062	[hypothetical yeast protein 1 [Mycoplasma capricolum S23] (fragment 18578-186578) (hypothetical protein - Mycoplasma capricolum S23)]	79	60	996

TABLE 2

EP 0 786 519 A2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start ID (nt)	Stop ID (nt)	match accession	match gene name	% sim	% ident	length (nt)
352	6	10106	8886	[gi 216834	[P47K <i>Pseudomonas chlororaphis</i>	79	59	1221
412	1	1153	578	[gi 113177	[putative <i>Bacillus subtilis</i>]	79	51	576
401	3	621	1124	[gi 706163	[Ribosomal Protein L10 (<i>Bacillus subtilis</i>)	79	66	504
516	1	702	352	[gi 805090	[NisA <i>Lactococcus lactis</i>	79	48	351
525	2	2457	1426	[gi 103371	[phosphotransferase synthetase (PDR-N) (<i>Bacillus subtilis</i>)	79	61	1032
					[PirH23326 <i>ASBCC</i> phosphoribosylformylglycinamide cyclo-ligase EC 6.3.3.1] - <i>Bacillus subtilis</i>			
530	4	2448	2825	[gi 1170707	[orf6 <i>Lactococcus sakei</i>	79	67	624
570	1	2	421	[gi 461600	[arginine permease substrate-binding subunit <i>Listeria monocytogenes</i>]	79	61	420
645	6	2663	3241	[gi 153888	[transport protein (<i>Salmonella typhimurium</i>)	79	62	579
683	1	75	174	[gi 1064795	[function unknown (<i>Bacillus subtilis</i>)	79	62	300
816	3	4700	3987	[gi 1107784	[orf-: novel antigen (<i>Staphylococcus aureus</i>)	79	62	714
2929	1	3	401	[gi 1124397	[glycine betaine transporter OpdU (<i>Bacillus subtilis</i>)	79	61	399
2937	1	357	202	[pir 52915 5539	[nitrate reductase alpha chain - <i>Bacillus subtilis</i> (fragment)	79	58	156
2940	1	768	385	[gi 119439	[putative (<i>Lactococcus lactis</i>)	79	72	384
2946	1	570	286	[gi 113267	[2-oxoglutarate dehydrogenase (adhA; EC 1.2.4.2) (<i>Bacillus subtilis</i>)	79	61	265
2959	1	3	212	[gi 170020	[nitrite reductase (nrR) (<i>Bacillus subtilis</i>)	79	59	210
3022	1	514	332	[gi 140686	[3-phosphoglycerate kinase (<i>Thermotoga maritime</i>)	79	61	183
3064	1	3	314	[gi 1104436	[pyruvate formate-lyase (<i>Haemophilus influenzae</i>)	79	60	312
3083	1	2	220	[gi 1149662	[hypD gene product (<i>Clostridium perfringens</i>)	79	56	219
3124	1	701	411	[gi 1139950	[large subunit of NADH-dependent glutamate synthase (<i>Plactonema boryanum</i>)	79	55	291
3181	1	607	326	[gi 1139950	[large subunit of NADH-dependent glutamate synthase (<i>Plactonema boryanum</i>)	79	59	282
3305	1	1	476	[gi 187784	[Cip-like ATP-dependent protease binding subunit (<i>Bos taurus</i>)	79	63	474
3718	1	536	270	[trk 2168891 C36	[lasc protein, inactive - <i>Lactococcus lactis</i> subsp. <i>lactis</i> (strain IL1403)	79	71	267
3724	2	159	401	[gi 1069366	[respiratory nitrate reductase (<i>Bacillus subtilis</i>)	79	64	243
3836	1	608	212	[gi 1524193	[Unknown (<i>Mycobacterium tuberculosis</i>)	79	65	297
3941	1	2	334	[gi 1419855	[deoxyribose aldolase (<i>Hycoplasma hoisensis</i>)	79	54	333
4113	1	3	341	[gi 113015	[gluconate kinase (<i>Bacillus subtilis</i>)	79	63	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (In)	Stop (In)	match accession	match gene name	sim	ident	length (nt)
4501	1	406	209	[gi 11021726	Unknown [Staphylococcus haemolyticus]	79	66	198
4612	1	2	238	[gi 460689	lrvC [Thermactomyces vulgaris]	79	58	237
2	1	2	1213	[gi 520753	DNA topoisomerase I [Bacillus subtilis]	78	64	1212
8	2	2266	1220	[gi 216151	DNA polymerase (gene L; tgg start codon) [Bacteriophage SP02] d[gi 579397	78	72	1047
					SP02 DNA polymerase (aa 1-648) [Bacteriophage SP02] Dif[A2198]QJBP52 DNA-directed DNA polymerase (EC 2.7.7.7) - phage P02			
9	12	1340	1089	[gi 1044787	function unknown [Bacillus subtilis]	78	57	252
12	6	6803	7702	[gi 116974	[NAD]-dependent NAD synthetase [Escherichia coli]	78	63	900
36	4	2941	3138	[gi 205053	glutamate permease [Escherichia coli]	78	53	198
53	115	17684	16221	[gi 1103941	YqIV [Bacillus subtilis]	78	58	1464
57	114	110520	12067	[gi 1102418	[lca] gene product [Staphylococcus carnosus]	78	65	1548
66	7	6798	5612	[gi 11212729	[YqIV] [Bacillus subtilis]	78	67	987
67	4	4029	4376	[gi 466612	[lka] [Escherichia coli]	78	71	348
91	9	110058	10942	[gi 467380	stage 0 sporulation [Bacillus subtilis]	78	50	885
102	12	4574	10130	[gi 1149426	putative [lactococcus lactis]	78	61	1557
112	6	3540	4463	[gi 185234	[cyc] gene product [Klebsiella oxytoca]	78	56	924
124	2	1888	1061	[gi 405622	Unknown [Bacillus subtilis]	78	60	826
130	3	1805	2260	[gi 11256616	putative [Bacillus subtilis]	78	71	456
133	1	751	377	[gi 1168060	[lmb] [Escherichia coli]	78	59	375
166	4	7125	6163	[gi 451216	[mannosephosphate isomerase [Streptococcus mutans]	78	63	963
186	1	1586	795	[gi 1128284	lcytanyl-rRNA synthetase [Bacillus subtilis]	78	63	792
195	4	2749	2315	[gi 11353874	Unknown [Rhodobacter capsulatus]	78	58	435
199	3	4279	3623	[gi 143525	succinate dehydrogenase cytochrome b558 subunit [Bacillus subtilis] d[gi 294410BSSC succinate dehydrogenase (EC 1.3.99.11) cytochrome 558 -	78	57	657
					protein [Bacillus subtilis sp P14931] [UVIC-BACSU EXCINUCLEASE ABC SUBUNIT C]			
223	3	1831	3523	[gi 139596	[Escherichia coli IS200 insertion sequence from EC063, partial], one product [Escherichia coli]	78	47	109

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	orf ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
299	4	1865	2149	gi 467439	temperature sensitive cell division [Bacillus subtilis]	78	62	283
321	9	7730	7315	gi 142579	[ORF 1 is homologous to an ORF downstream of the spor gene of E.coli; RF3]	78	55	426
352	4	3716	3944	gi 139060	lectin 1 [Pneumocytella carinii]	78	42	231
352	5	7592	6093	gi 903587	[NADH dehydrogenase subunit 5 [Bacillus subtilis] AP P39755 NDHP-BAC5U NADH DEHYDROGENASE SUBUNIT 5 (EC 1.6.5.3) NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5]	78	58	1500
376	1	2	583	gi 551633	[dithiobiotin synthase [Bacillus sphaericus]]	78	34	582
424	2	1595	1768	gi 1524117	[alpha-acetolactate decarboxylase [Lactococcus lactis]]	78	68	174
450	1	1914	948	gi 1030668	[NAD(P)H oxidoreductase, isoflavone reductase homologue [Selaginella tuberosa]]	78	63	927
558	1	762	1562	gi 1511588	[bisfunctional protein [Methanococcus jannaschii]]	78	60	201
670	3	1152	1589	gi 11122759	[unknown [Bacillus subtilis]]	78	64	438
714	1	64	732	gi 1143460	[17 kd minor sigma factor [rp0F, sigma-17tgc start codon] [Bacillus subtilis]]	78	57	669
814	1	3	168	gi 1377853	[unknown [Bacillus subtilis]]	78	59	366
981	1	1381	692	gi 143802	[GcrC2 [Bacillus subtilis]]	78	64	690
995	2	978	1277	gi 1296947	[uridine kinase [Escherichia coli]]	78	64	252
1045	1	3	401	gi 1407784	[orf-1; novel antigen [Staphylococcus aureus]]	78	61	399
1163	2	368	186	gi 410117	[lumionopimelate decarboxylase [Bacillus subtilis]]	78	54	183
2191	1	294	399	gi 1215098	[excisionase [Bacteriophage 154a]]	78	65	396
2933	1	2	181	gi 1204436	[pyruvate formate-lyase [Haemophilus influenzae]]	78	73	180
3041	2	1129	317	gi 1624632	[GII.1 [Escherichia coli]]	78	53	189
3581	1	105	401	gi 1763186	[3-ketocacyl-CoA thiolase [Saccharomyces cerevisiae]]	78	55	297
3709	1	3	230	gi 460689	[rncG [Thermostaphylococcus vulgaris]]	78	58	228
3974	1	528	1265	gi 1558039	[unknown [Bacillus subtilis]]	78	65	264
3980	1	3	401	gi 139956	[110lc [Bacillus subtilis]]	78	62	399
4056	1	647	334	gi 1256615	[dihydroxy-acid dehydratase [Bacillus subtilis]]	78	55	294
4114	1	630	316	gi 5093721503	[hypothetical protein - Trypanosoma brucei]	78	62	315
4185	1	3	179	gi 133950	[large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]]	78	58	177

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4235	1	655	329	911588839	[Unknown [Bacillus subtilis]]	78	60	327
4352	1	541	302	911603768	[PutI protein, indoleamine-5-propionate hydrolase [Bacillus subtilis]] 911603768 [PutI protein, indololone-5-propionate hydrolase [Bacillus subtilis]]	78	63	240
4368	1	612	307	9111353678	[heavy-metal transporting P-type ATPase 1 [Proteus mirabilis]]	78	59	306
4461	1	428	216	911176641	[glutamate synthase (GOGAT) [Pseudomonas putida]]	78	36	213
4510	1	474	238	91119956	[IGLC [Bacillus subtilis]]	78	65	237
3	2	2969	2073	91110984	[PROV [Bacillus subtilis]]	77	56	897
12	2	2426	1965	911467315	[ribosomal protein L9 [Bacillus subtilis]]	77	59	462
27	1	2	388	9111212228	[YgbI [Bacillus subtilis]]	77	63	387
39	2	590	1252	911140054	[phenylalanyl-tRNA synthetase beta subunit [AA 1-304] [Bacillus brevis]]	77	60	663
42	6	2704	2931	9111606241	[30S ribosomal subunit protein S14 [Escherichia coli] sp P02370 RS14_ECOLI 10S RIBOSOMAL PROTEIN S14 [ISUB 2-101]]	77	65	228
46	16	1459	16622	911237759	[mitochondrial formate dehydrogenase precursor [Solanum tuberosum] p1113Q1272 [Q1272 formate dehydrogenase (EC 1.2.1.12) precursor, mitochondrial - potato]]	77	55	1164
100	4	4562	4002	9111240128	[lmpF1 [Staphylococcus aureus]]	77	54	561
102	8	5378	5713	9111311682	[acetolactate synthase (Thermus aquaticus)]	77	57	336
109	7	4742	5383	9111210637	[Unknown [Bacillus subtilis]]	77	56	642
117	1	2	1228	91112337015	[ORF4 [Bacillus subtilis]]	77	53	1227
124	10	6323	7688	9111405819	[thymidine kinase [Bacillus subtilis]]	77	61	636
147	3	1146	985	9111849027	[hypothetical 15.9-kDa protein [Bacillus subtilis]]	77	37	162
152	10	7354	7953	9111205583	[spermidine/polyarginine transport ATP-binding protein [Haemophilus influenzae]]	77	55	600
169	2	1004	1282	9111473825	[elongation factor EF-Ts [Escherichia coli]]	77	58	279
184	2	180	1147	911216314	[esterase [Bacillus stearothermophilus]]	77	60	768
189	7	3296	3868	9111653809	[ORF3 [Clostridium perfringens]]	77	48	573
193	1	132	290	9111303768	[YqeH [Bacillus subtilis]]	77	54	159
195	16	8740	8114	9111499520	[M. jannaschii predicted coding region M0798 [Methanococcus jannaschii]]	77	44	327
205	6	5428	5204	9111216340	[ORF for adenylylate kinase [Bacillus subtilis]]	77	61	225

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

5
10
15
20
25
30
35
40
45
50

conig	orf ID	start (nt)	stop (nt)	match	gene name	sim	ident	length (nt)
205	39	14795	11502	911786155	[Ribosomal Protein L23 [Bacillus subtilis]]	77	62	294
211	5	1908	12084	911410132	[ORF8 [Bacillus subtilis]]	77	47	177
217	5	3478	14416	911496254	[Fibronectin/fibrinogen-binding protein [Streptococcus pyogenes]]	77	54	939
232	1	267	1998	911407784	[orf-1; novel antigen [Staphylococcus aureus]]	77	57	722
233	2	1819	11446	911461408	[Unknown [Bacillus subtilis]]	77	61	474
243	3	2661	2299	911516155	[unconventional myosin [Sus scrofa]]	77	32	363
299	1	68	769	911467436	[Unknown [Bacillus subtilis]]	77	54	702
301	4	1468	1283	911950071	[AMP-bind. pyridine kinase [Hycoplasma capricolum] pir S48605 S48605 [fragment]]	77	48	186
302	5	2741	3211	911509980	[phab [Bacillus subtilis]]	77	57	471
302	7	1835	4863	911477783	[ruvB protein [Escherichia coli]]	77	60	1029
302	9	5402	4797	9111070015	[protein-dependent [Bacillus subtilis]]	77	60	606
312	1	99	1391	91143165	[malic enzyme [EC 1.1.1.38] [Bacillus stearothermophilus] pir A33307 DEASX5 [fragment]]	77	62	1293
312	2	1541	2443	911139955	[carboxyltransferase beta subunit [Synechococcus PCC7942]]	77	58	903
321	5	5666	4596	91139444	[fumurate (ftrc) [aa 1-462] [Bacillus subtilis]]	77	65	1071
354	1	47	564	911154634	[YaaH [Bacillus subtilis]]	77	57	522
365	1	2	1021	911143374	[phosphotriose glycinamide synthetase (PUR-D; gtc start codon) [Bacillus subtilis]]	77	62	1020
374	1	1	708	9111405446	[transketolase [Bacillus subtilis]]	77	61	708
385	1	1128	565	9111533093	[endonuclease III [Bacillus subtilis]]	77	63	564
392	2	594	1940	911556014	[UDR-N-acetyl muramate-alanine ligase (Bacillus subtilis) (EC 3.2.8) (sp P40778 IMURC_BACSU UDP-N-ACETYLURIDYL-L-ALANINE SYNTHETASE) (FRAGMENT)]	77	65	1347
405	5	4079	1370	911153912	[YgbW [Bacillus subtilis]]	77	64	510
487	4	1302	1472	911432427	[orf1 gene product [Actinobacter calcoaceticus]]	77	48	171
522	1	2	562	pir A01179 SYNS	[tyrosine-tRNA ligase (EC 6.1.1) - Bacillus stearothermophilus]	77	63	561

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% silent	length (nt)
523	2	1587	1351	91 11387979	1448 Identity over 302 residues with hypothetical protein from <i>Synechocystis</i> sp. accession D6006-2B; expression induced by environmental stress; some similarity to Glyceral transferases; two potential membrane spanning helices [Bacillus subtilis]	77	48	237
536	2	983	612	91 113366	adenylosuccinate lyase (PUR-B) [Bacillus subtilis] pir C29126 NB2SDS	77	61	372
548	2	339	872	91 1431387	espartate transcarboxylase [Bacillus subtilis]	77	56	534
597	1	2	481	91 190498	hypothetical protein [Bacillus subtilis]	77	13	480
631	2	1747	1313	91 1387377	[ORFIA] [Bacillus subtilis]	77	64	415
642	1	85	360	91 146971	epip gene product (Staphylococcus epidermidis)	77	61	276
659	1	125	1219	91 1072381	glutamyl-aminopeptidase (Lactococcus lactis)	77	62	1095
670	4	1587	1820	91 1122760	lunknem [Bacillus subtilis]	77	58	234
789	1	2	391	91 1377823	laminopeptidase [Bacillus subtilis]	77	65	390
815	1	10	573	91 1303861	[YggN] [Bacillus subtilis]	77	49	564
1079	1	1	225	91 120144	II. influenzae predicted coding region II.0594 (Influenphilus influenzae)	77	55	225
1083	1	3	188	91 1460928	18969 [Saccharomyces cerevisiae]	77	66	186
1942	1	415	209	91 160047	[p10] acidic basic repeat antigen (Plasmodium falciparum) D111292221 A29332	77	38	207
2559	1	1	171	91 1498034	[H. jannaschii] predicted coding region HJ0155 (Methanococcus jannaschii)	77	61	171
2933	2	243	401	91 142370	pyruvate formate-lyase (AA 1-760) [Escherichia coli] pir S01768 S01788	77	72	159
2966	1	56	292	91 1524397	formate C-acetyltransferase (EC 2.3.1.54) - <i>Escherichia coli</i>	77	45	237
2976	1	614	309	91 140003	glycine betaine transporter QMD (Bacillus subtilis)	77	60	306
					oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] P P23129 P001-BACSU	77	55	225
					[ALPHA-KETODIURATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2)]	77	63	165
					spore germination and vegetative growth protein [Bacillus subtilis] 78	77	56	387
2988	1	601	377	91 138665	Probable operon with off. Possible alternative initiation codon, <i>aaa</i> 2151-2153. Homology with acetyltransferases; putative <i>Bacillus subtilis</i>	77	61	279
2990	1	331	167	91 142562	ATP synthase epsilon subunit [Bacillus megaterium] pir B26599 PMBEM H ⁺ -transporting ATP synthase [EC 3.6.1.31] psilon chain - <i>Bacillus megaterium</i>	77	63	165
3032	1	3	389	91 1468330	alcohol dehydrogenase 2 [Entamoeba histolytica]	77	50	195
3057	1	1	195	91 1468764	Mock gene product [Rhizobium meliloti]	77	56	387

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

5
10
15
20
25
30
35
40
45
50
55

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4008	1	726	400	gi 603768	[uri protein, imidazolone-5-propanolate hydrolase (Bacillus subtilis) gi 603766 urt protein, imidazolone-5-propanolate hydrolase Bacillus subtilis]	77	52	327
4008	1	703	386	gi 216278	[branched S synthetase I (Bacillus brevis)]	77	55	318
4110	1	3	368	pir 652915 S5329	[lactate reductase alpha chain - Bacillus subtilis (fragment)]	77	61	366
4115	1	1	348	gi 5157205	[67 kDa Myosin-crossreactive streptococcal antigen (Streptococcus yogenesi)]	77	65	348
4225	1	590	297	gi 1322245	[nevalonate pyrophosphate decarboxylase (Rattus norvegicus)]	77	60	294
4611	2	494	127	gi 1508979	[cRP-binding protein (Bacillus subtilis)]	77	57	168
4668	1	361	182	pir 652915 S5329	[lactate reductase alpha chain - Bacillus subtilis (fragment)]	77	61	180
25	1	2	1627	gi 1150620	[HmSA (Streptococcus pneumoniae)]	76	58	1626
38	5	1488	2537	pir 643577 A335	[regulatory protein pirR - Clostridium perfringens]	76	57	1050
52	5	2962	4041	gi 11161061	[dixygenase (Methylobacterium extorquens)]	76	62	1080
56	20	27389	127955	gi 1467402	[unknown (Bacillus subtilis)]	76	56	587
57	15	12046	13229	gi 1206040	[weak similarity to karsin (Caenorhabditis elegans)]	76	40	174
91	2	1062	2261	gi 475715	[acetyl coenzyme A acetyltransferase (Thiobacillus octobutylicum)]	76	57	1200
94	2	618	1624	gi 1467432	[unknown (Bacillus subtilis)]	76	62	807
94	5	2965	3228	gi 897793	[y88 gene product (Padiococcus acidilactic)]	76	52	264
98	8	5922	6326	gi 464427	[methionyl-tRNA synthetase (Bacillus subtilis)]	76	53	405
104	3	1322	1885	gi 1216151	[DNA polymerase (gene L, rts start codon) Bacteriophage SP021 gi 1579197 SP02 DNA polymerase (aa 1-648) (Bacteriophage SP02) pir A11498 DJBPS2 DNA-directed DNA polymerase (EC 2.7.7.7) - phage P02]	76	63	564
124	9	8114	7055	gi 853776	[peptide chain release factor 1 - Bacillus subtilis pir S55437 S55437 peptide chain release factor 1 - Bacillus subtilis]	76	58	1080
164	5	2832	3311	gi 1264976	[prolyl-tRNA synthetase (Haemophilus influenzae)]	76	53	480
168	2	2617	1841	gi 1177253	[putative ATP-binding protein of ABC-type (Bacillus subtilis)]	76	58	177
189	2	163	888	gi 647384	[unknown (Bacillus subtilis)]	76	63	726
235	3	2353	3518	gi 143936	[folyl-polyglutamate synthetase (Bacillus subtilis) pir B80661 B40646 folC - Bacillus subtilis]	76	53	1266
236	1	325	925	gi 1146197	[putative (Bacillus subtilis)]	76	54	591
237	8	5223	5241	gi 1219281	[F11G3.6 (Caenorhabditis elegans)]	76	47	219

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
263	5	5490	4585	[gi 1510148]	1,4hydroxipicolinate synthase [Methanococcus jannaschii]	76	49	906
304	3	1051	1794	[gi 666382]	putative membrane spanning subunit [Bacillus subtilis] [gi 552187 852282]	76	60	744
312	4	3611	4624	[gi 113312]	probable membrane spanning protein - Bacillus subtilis	76	56	1014
343	1	2	1036	[gi 405956]	[6-phospho-1-fructokinase (gtg start codon: EC 2.7.1.11) [Bacillus teethermophilus]]	76	59*	1035
347	1	409	1701	[gi 1396304]	[yeaE [Escherichia coli]]	76	72	1293
358	1	672	1907	[gi 1146215]	[39.01 Identity to the Escherichia coli S1 ribosomal protein: putative [Bacillus subtilis]]	76	58	1236
371	1	1	222	[gi 537084]	[alternate gene name: mgf; CG Site No. 497 [Escherichia coli]]	76	61	222
379	4	4331	4658	[gi 113268]	[dihydrofolate reductase [Escherichia coli]]	76	61	328
404	5	4022	4492	[gi 1103823]	[dihydrofolate reductase transacetylase (tdhB; EC 2.3.1.61) [Bacillus subtilis]]	76	60	471
411	1	2	307	[gi 36025]	[orf YKL027w [Saccharomyces cerevisiae]]	76	55	306
412	3	4356	2054	[gi 105664]	[hst [Bacillus subtilis]]	76	57	1503
546	1	273	995	[gi 133821]	[streptococcal pyrogenic exotoxin type C (spec) precursor Streptococcus pyogenes]	76	36	723
588	1	1054	557	[gi 1005220]	[nutS [Bacillus subtilis]]	76	61	496
591	1	16	735	[gi 885934]	[tpS [Synchococcus sp.]	76	44	720
602	2	175	798	[gi 1486222]	[tpG homologue [Rhizobium sp.]	76	52	624
619	2	547	290	[gi 110611]	major capsid protein [Human cytomegalovirus]	76	47	258
660	4	2568	3302	[gi 904199]	[hypothetical protein [Bacillus subtilis]]	76	55	735
677	1	452	228	[gi 40177]	[spooF gene product [Bacillus subtilis]]	76	58	225
962	1	24	206	[gi 12443]	[adenylosuccinate synthetase [Bacillus subtilis] [gi 29726 PUR_A_BCSU]]	76	67	183
978	1	1198	580	[gi 151333]	[M. jannaschii predicted coding region Mj1322 [Methanococcus jannaschii]]	76	56	579
997	1	486	244	[gi 1467154]	[No definition line found [Mycobacterium leprae]]	76	38	243
1563	1	529	266	[gi 1103984]	[tpG [Bacillus subtilis]]	76	52	264
2184	1	361	182	[gi 566706]	[capJ [Staphylococcus aureus]]	76	38	180
2572	1	1	387	[gi 113898]	[transport protein [Salmonella typhimurium]]	76	65	387

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (Int)	Stop (Int)	match accession	match gene name	% sim	% ident	length (Int)
2942	1	29	400	[gi 710020	nitrite reductase [nrB] (Bacillus subtilis)	76	59	372
2957	1	377	216	[gi 151251	hypothetical protein (SP:Pa204) [Methanococcus jannaschii]	76	47	162
2980	1	554	279	[gi 1105464	[ATST (Bacillus subtilis)]	76	53	276
3015	1	649	326	[gi 1408115	ornithine acetyltransferase (Bacillus subtilis)	76	61	324
3124	1	13	174	[gi 1802705	[ORF o401] (Escherichia coli)	76	65'	162
3179	1	3	161	[gi 168477	[arcE] (arcE-dependent glutamate synthase [Zea mays] [pir A3896 A3896	76	53	159
3769	1	2	379	[gi 39956	glutamate synthase (ferredoxin) (EC 1.4.7.1) - a zinc [IGIC (Bacillus subtilis)]	76	55	378
1402	1	3	314	[gi 1510398	[terz] (pyochelin binding protein [Methanococcus jannaschii])	76	52	312
3928	1	798	400	[gi 143016	[paromomycin (Bacillus subtilis)]	76	59	399
4159	1	757	386	[sp P0544 HRSP]	[HEMUSILIN-RESISTANT SURFACE PROTEIN (FRAGMENTS)]	76	66	372
4204	1	17	331	[gi 296464	[ATPase (Lactococcus lactis)]	76	56	315
4394	1	494	129	[gi 1987255	[Henkes disease gene (Homo sapiens)]	76	48	216
4506	1	2	313	[gi 216746	[D-lactate dehydrogenase [Lactobacillus plantarum]]	76	47	312
4546	1	477	1247	[gi 1133950	[large subunit of RASH-dependent glutamate synthase [Pleotomella boryanum]]	76	61	231
4596	1	379	191	[gi 560037	[cellulose synthase [Acetobacter xylinum]]	76	70	189
4	5	5257	4337	[gi 1802512	[ORF o29] (Escherichia coli)	75	59	931
6	1	164	952	[gi 140960	[orfase (Escherichia coli)]	75	56	789
12	1	5935	1944	[gi 1467336	[unknown (Bacillus subtilis)]	75	57	1932
23	18	18272	17310	[gi 129613]	[0-acetylserine sulfhydrylase B (Methylophilus autotrophicus)]	75	55	963
25	3	2356	3393	[gi 1520419	[PksX (Bacillus subtilis)]	75	56	1038
26	8	5765	6037	[gi 1256517	[unknown (Schizosaccharomyces pombe)]	75	45	273
46	13	11186	12058	[gi 148972	[nitrate transporter (Synechococcus sp.)]	75	46	873
51	7	3474	3677	[gi 1115607	[sporulation protein (Bacillus subtilis)]	75	61	204
53	16	16850	16590	[gi 1143402	[recombination protein (ttg start codon), (Bacillus subtilis) gi 1103923 RecN	75	51	261
74	1	3572	2568	[gi 1204847	[ornithine carbamoyltransferase (Mamophilus influenzae)]	75	61	1005

TABLE 2

S. durus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start Int	Stop Int	match accession	match gene name	% sim	% ident	length (nt)
83	3	4628	3930	[gi 143368	phosphoribosyl glycaminidine synthetase I (PUR-1; gtc start odon)	75	63	699
85	5	5588	4878	[gi 143367	phosphoribosyl aminodarole succinocboxamide synthetase (PUR-C; gtc start odon)	75	55	711
85	6	6625	7530	[gi 1503916	[codon] (Bacillus subtilis)	75	53	906
87	3	2340	3590	[gi 1064813	[YgtA (Bacillus subtilis)]	75	53	1251
87	6	6084	6896	[gi 1064810	[homologous to sp:PHOR_BACSU (Bacillus subtilis)]	75	56	900
108	2	1844	1503	[gi 1001624	[function unknown (Bacillus subtilis)]	75	61	813
110	3	1748	13727	[gi 1147593	[hypothetical protein (Synchocystis sp.)]	75	51	342
110	7	4151	5252	[gi 1117251	[putative ppGpp synthetase (Streptomyces coelicolor)]	75	55	1980
120	14	11266	10649	[gi 1524394	[cluD gene product (Bacillus subtilis)]	75	75	900
121	5	2050	4221	[gi 1154632	[ORF-2 upstream of gbaB operon (Bacillus subtilis)]	75	55	618
124	1	283	143	[gi 105622	[NrdE (Bacillus subtilis)]	75	54	2122
128	1	81	1139	[gi 143316	[unknown (Bacillus subtilis)]	75	56	141
130	8	5760	5903	[gi 1256554	[hypothetical gene products (Bacillus megaterium)]	75	48	1059
136	2	4480	3185	[gi 1467103	[5'-AU identity with Neisseria gonorrhoeae regulatory protein NRD; putative (Bacillus subtilis)]	75	67	144
161	10	5439	5798	[gi 1001195	[lucy1]-tRNA synthetase (Bacillus subtilis)]	75	54	1296
172	4	3819	2995	[gi 1755153	[hypothetical protein (Synchocystis sp.)]	75	55	360
179	1	2024	1107	[gi 143037	[ATP-binding protein (Bacillus subtilis)]	75	52	825
195	10	9528	9174	[sp P25745 YCP_	[porphobilinogen deaminase (Bacillus subtilis)]	75	58	918
200	4	2605	4596	[gi 142440	[hypothetical protein in PUR-5 REGION (ORF-15) (FRAGMENT). (Bacillus subtilis)]	75	60	156
206	3	6900	5620	[gi 1256135	[ATP-dependent nuclease (Bacillus subtilis)]	75	56	1922
216	2	159	389	[gi 1052400	[unknown (Schizosaccharomyces pombe)]	75	53	1281
229	1	29	847	[gi 1205958	[branched chain aa transport system II carrier protein (Haemophilus influenzae)]	75	49	819
230	2	518	1714	[gi 971337	[nitrite extrusion protein (Bacillus subtilis)]	75	53	1197
231	1	2240	1122	[gi 1002521	[HutC (Bacillus subtilis)]	75	54	1119
233	3	3314	1859	[gi 167405	[unknown (Bacillus subtilis)]	75	59	546

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	orf_id	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
269	1	325	164	[gi 1511246]	methyl coenzyme M reductase system, component A2 [Methanococcus jannaschii]	75	50	162
282	1	1389	772	[gi 1511604]	[M. jannaschii] predicted coding region MJ1651 [Methanococcus jannaschii]	75	46	618
304	4	1773	2263	[gi 17205228]	surfacein [Haemophilus influenzae]	75	55	489
312	3	2437	3387	[gi 185621]	undefined open reading frame [Bacillus stearothermophilus]	75	62	951
312	5	4622	6403	[gi 1041097]	[Pyruvate kinase] [Bacillus psychrophilus]	75	57	1782
319	1	353	877	[gi 1212728]	[Qhi] [Bacillus subtilis]	75	54	525
320	5	4321	5031	[gi 1070361]	[OMP decarboxylase] [Lactococcus lactis]	75	56	711
320	6	5010	5632	[gi 143394]	[OMP-PRPP transferase] [Bacillus subtilis]	75	60	633
337	4	1519	2088	[gi 148733]	[lactate synthase II] [Bacillus subtilis]	75	58	570
394	2	669	1271	[gi 304976]	[matchs PS00017: ATP-GTP_A and PS00310: EFACTOR_GTP; similar to longation factor G, Teff/Pefo tetraacycline-resistance proteins [Escherichia coli]	75	51	603
423	1	127	570	[gi 118319]	[unknown] [Pseudomonas aeruginosa]	75	59	444
433	2	1603	1929	[gi 149211]	[lactolactate synthase] [Klebsiella pneumoniae]	75	63	327
446	2	176	1540	[gi 12441]	[dihydroorotate] [Bacillus caldolyticus]	75	62	1365
466	1	494	249	[gi 1149682]	[potF gene product] [Clostridium perfringens]	75	55	246
496	1	3	794	[gi 143582]	[spolIIEA protein] [Bacillus subtilis]	75	59	792
498	2	824	1504	[gi 143328]	[phop protein (put.); putative] [Bacillus subtilis]	75	47	681
499	2	1051	1634	[gi 11387979]	448 identity over 302 residues with hypothetical protein from <i>Synchocystis</i> sp. accession DK4006_CD; expression induced by environmental stress; none similarity to glycosyl transferases; two potential membrane-spanning helices [Bacillus subtilis]	75	51	564
568	1	641	453	[pri]cc110]c41	tricetylglycerol lipase (Ec 3.1.1.3) 2 - [Hypoclesma mycoides subsp. mycoides] (SSC1)	75	50	189
613	2	430	233	[gi 330993]	[legumain protein] [Salmarina herpesvirus 2]	75	75	198
621	1	1	525	[gi 239754]	[spec] [Streptococcus pyogenes]	75	43	525
632	5	1409	2474	[gi 1176401]	[EpiG] [Staphylococcus epidermidis]	75	51	666
646	2	454	657	[gi 172442]	[ribonuclease P] [Saccharomyces cerevisiae]	75	37	204
657	1	3	347	[gi 182541]	[Jorf_02316] [Escherichia coli]	75	47	345
750	1	1662	832	[gi 16971]	[epip gene product] [Staphylococcus epidermidis]	75	57	631

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (Int)	Stop (Int)	match accession	match gene name	length (nt)
1	1	2	481	[gi 1303901]	[Yqnt (Bacillus subtilis)]	75
754	1	2	481	[gi 1303901]	[multidrug resistance protein (Haemophilus influenzae)]	75
761	1	563	593	[gi 120545]	[lactococcus lactis subsp. lactis (strain 11103)]	75
775	1	951	482	[pi 1B26889 B368]	[lacto protein, inactive - Lactococcus lactis subsp. lactis (strain 11103)]	75
783	1	1	180	[gi 143316]	[gap gene products (Bacillus megaterium)]	75
800	1	318	160	[gi 1509411]	[nfra protein (Asporhizobium caulinodans)]	75
811	1	1117	560	[gi 141434]	[Rho Factor (Bacillus subtilis)]	75
940	1	493	329	[gi 1276985]	[arginase (Bacillus caldovelox)]	75
971	1	2	37	[gi 1001373]	[hypothetical protein (Synechocystis sp.)]	75
1059	1	184	232	[gi 1756480]	[L-glutamin-D-fructose-6-phosphate amidotransferase (Bacillus subtilis)]	75
1109	2	219	374	[gi 143331]	[alkaline phosphatase regulatory protein (Bacillus subtilis)]	75
					[pir 212650 B27650 regulatory protein phoR - Bacillus subtilis sp P23451 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN NOR (EC 2.7.3.-)]	75
1268	1	271	137	[gi 1704135]	[ornithine acetyltransferase (Bacillus stearothermophilus)]	75
					[sp Q10708 ANG1_BACSU GLUTAMATE-N-ACETYLTRANSFERASE (EC 2.3.1.35) ORNITHINE ACYLTRANSFERASE (EC 2.3.1.20) / (OFTASE) / HIND-ACID ACETYLTRANSFERASE (EC 2.3.1.1) (N-ACETYLGLUTAMATE INTIA)]	75
1500	1	324	163	[gi 1205488]	[excinucllease ABC subunit B (Haemophilus influenzae)]	75
1520	1	796	400	[gi 1002521]	[Mott (Bacillus subtilis)]	75
1510	1	770	187	[gi 1204435]	[pyruvate formate-lyase activating enzyme (Haemophilus influenzae)]	75
3105	1	1	180	[gi 104107]	[Pyruvate kinase (Bacillus psychrophilus)]	75
3117	1	45	212	[gi 899317]	[peptide synthetase module [Microcytis aeruginosa] ptk S9111 S4911] (probable amino acid activating domain - <i>icrocytis aeruginosa</i> (fragment) (SIB 14-4-526)]	75
3139	2	139	345	[gi 145294]	[adenine phosphoribosyl-transferase (Escherichia coli)]	75
3860	1	618	310	[gi 1009366]	[respiratory nitrate reductase (Bacillus subtilis)]	75
3911	1	48	401	[gi 133991]	[ATP synthase subunit beta (Bacillus subtilis)]	75
3957	1	2	379	[pi 16889 B368]	[l-isopropylmalate dehydrogenase (EC 4.2.1.33) chain leuC - <i>lactococcus lactis</i> subsp. <i>lactis</i> (strain 111403)]	75
4005	1	5	259	[gi 16746]	[D-lactate dehydrogenase (Lactobacillus plantarum)]	75
4080	1	73	333	[gi 150555]	[deoxyribose aldolase (Hycoplasma hominis)]	75

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. entericus - putative coding regions of novel proteins similar to known proteins

Contig	Olf ID	Start	Stop	match	match gene name	% sim	% ident	length (nt)
	ID	[Int]	[Int]	cassion				
4111	1	1	1	139	[gi 1149435] putative [lactococcus lactis]	75	57	119
4136	1	602	303	[gi 450688]	[gi 309629] putative [Escherichia coli] protein - Escherichia coli pir [gi 309629] hypothetical protein A - Escherichia coli (SUB 40-520)	75	56	300
4144	1	668	336	[gi 48972]	[gi 309629] putative transporter [Synchococcus sp. 1]	75	49	333
4237	1	664	1374	[gi 1399950]	large subunit of NADH-dependent glutamate synthase [Pleotomella horvatum]	75	55	291
4306	1	73	318	[gi 29260]	major surface glycoprotein [Pneumocystis carinii]	75	68	246
4343	1	715	359	[gi 1204632]	methylated-DNA-protein-cysteine methyltransferase [Haemophilus influenzae]	75	52	357
4552	1	1	620	[gi 294464]	[ATPase] [lactococcus lactis]	75	55	109
38	9	5776	6126	[gi 143793]	[HnPC] [Escherichia coli]	74	50	331
50	8	6910	6321	[gi 129988]	hypothetical protein [Bacillus subtilis]	74	55	690
56	9	11070	12221	[gi 1000451]	[Trp] [Bacillus subtilis]	74	57	1452
64	2	3266	1622	[gi 41015]	[aspartate-tRNA ligase] [Escherichia coli]	74	57	645
66	1	6	5063	[gi 4848]	[Yqhu] [Bacillus subtilis]	74	47	216
67	118	14334	14897	[gi 1510631]	[endoglycanase] [Methanococcus jannaschii]	74	52	564
102	115	12561	13116	[gi 149429]	[putative] [lactococcus lactis]	74	67	576
102	116	13121	144159	[gi 149435]	[putative] [lactococcus lactis]	74	57	1299
108	4	4873	3902	[gi 19478]	ATP binding protein of transport ATPases [Bacillus firmus] ir [gi 5486] [S15486]	74	59	972
					ATP-binding protein - Bacillus firmus plp26946 [YKR_BACPT1] HYPOTHETICAL			
					ATP-BINDING TRANSPORT PROTEIN			
116	5	8574	17093	[gi 1205430]	dipeptide transport system permease protein [Haemophilus influenzae]	74	49	1482
120	7	4342	4803	[gi 1146970]	ribonucleoside triphosphate reductase [Escherichia coli] pir [gi 7331] [A47331]	74	58	462
					anabolic ribonucleotide reductase - Escherichia coli			
121	7	5961	6581	[gi 11107526]	[trg start] [Campylobacter coli]	74	51	621
128	13	2320	3531	[gi 143318]	[phosphoglycerate kinase] [Bacillus angolensis]	74	57	1212
130	7	5237	5791	[gi 1256653]	[Dna-binding protein] [Bacillus subtilis]	74	60	555
136	3	6745	5150	[gi 142076]	[histidase] [Bacillus subtilis]	74	58	1596
145	2	664	1368	[gi 107773]	[deva gene product] [Anabaena sp.]	74	45	705
152	1	552	277	[gi 1377833]	[unknown] [Bacillus subtilis]	74	54	276

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
164	1.0	11064	111375	0 1 580000	[ORF3 gene product (Bacillus subtilis)]	74	52	312
175	1.2	3109	2624	0 1 642556	[Unknown (Rhizobium meliloti)]	74	34	486
175	1.9	6064	5612	0 1 854656	[Na/H antiporter system ORP2 (Bacillus acetylphillus)]	74	46	453
195	1.1	11346	110339	0 1 1204410	[Hypothetical protein (Sp; P25745) (Haemophilus influenzae)]	74	55	1008
205	1.7	9619	9059	0 1 104979	[Ribosomal protein L6 (Bacillus subtilis)]	74	64	561
216	1.7	5574	6710	0 1 1146207	[Putative (Bacillus subtilis)]	74	63	1137
241	1.3	4521	3334	0 1 694121	[Malate thiokinase (Methylbacterium extorquens)]	74	52	1188
246	1.6	3305	2799	0 1 1467314	[Single strand DNA binding protein (Bacillus subtilis)]	74	64	507
249	1.4	6551	5313	0 1 1524397	[Lysine betaine transporter Opd (Bacillus subtilis)]	74	55	1239
261	1.7	4389	4081	0 1 809542	[CbrB protein (Lewisia chrysanthemum)]	74	42	309
278	1.6	5714	4665	0 1 1204872	[ATP-binding protein (Haemophilus influenzae)]	74	54	1050
309	1.1	1220	666	0 1 1205579	[Hypothetical protein (GB:U44003_3021) (Haemophilus influenzae)]	74	53	555
315	1.2	1473	862	0 1 1143398	[Quinol oxidase (Bacillus subtilis)]	74	57	612
320	1.1	1	1065	0 1 143389	[Glutaminase of carbamyl phosphate synthetase (Bacillus subtilis) (EC 6.3.5.5) (Bacillus subtilis)]	74	60	1065
405	1.2	1742	1311	0 1 1303915	[Yvh2 (Bacillus subtilis)]	74	65	432
431	1.5	2503	12270	0 1 473902	[Alpha-acetolactate synthase (Lactococcus lactis)]	74	56	768
452	1.1	1	942	0 1 413982	[Isp-SR gene product (Bacillus subtilis)]	74	52	942
461	1.1	3	1193	0 1 554494	[Homoserine dehydrogenase (Bacillus subtilis)]	74	51	1191
461	2	1174	1407	0 1 40211	[Threonine synthase (thrc) (AA 1-352) (Bacillus subtilis) tr A25364 A25364	74	56	234
462	1.2	402	734	0 1 42520	[Thioredoxin (Bacillus subtilis)]	74	62	333
478	1	574	370	0 1 1499005	[Glycyl-tRNA synthetase (Methanococcus jannaschii)]	74	52	255
501	2	739	1740	0 1 21040	[acid glycoprotein (Streptococcus pyogenes)]	74	58	1002
551	2	4083	2791	0 1 141040	[Glutamate-1-semialdehyde 2,1-aminotransferase (Bacillus subtilis) plr D42720 D02728 glutamate-1-semialdehyde 2,1-aminomutase (EC 4.3.3.6) - Bacillus subtilis]	74	51	1293

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (Int)	Stop (Int)	match accession	match gene name	% sim	% ident	length (nt)
573	1	1	477	[gi 1006605	[hypothetical protein [Smychocystis sp.]	74	45	477
596	12	1298	1298	[gi 1301853	[YQ9F [Bacillus subtilis]	74	55	483
618	2	2924	1758	[gi 116237	[21.4% of identity to trans-acting transcription factor of <i>Sacharomyces cerevisiae</i> ; 23% of identity to sucrose synthase of <i>Zea mays</i> ; putative [Bacillus subtilis]	74	55	1167
659	2	1269	1595	[gi 1072380	[ORF3 [Lactococcus lactis]	74	62	327
724	1	173	188	[gi 143374	[phosphoribosyl glycnamide synthetase (PUR-D; gta start codon) Bacillus subtilis]	74	58	186
743	2	604	1209	[gi 153833	[ORF1; putative [Streptococcus parasanguinis]	74	50	606
836	1	2	259	[gi 143458	[ORF V [Bacillus subtilis]	74	47	258
989	2	443	724	[gi 1303994	[YQ9H [Bacillus subtilis]	74	46	282
1106	1	1	492	[gi 146910	[lipid gene product [Staphylococcus epidermidis]	74	54	492
1135	2	373	528	[gi 141948	[lipA-2d gene product [Bacillus subtilis]	74	48	156
1234	1	417	412	[gi 495245	[lrcd gene product [Fusiformis chrysanthemi]	74	36	366
2566	1	2	236	[gi 1119701	[lbcC gene product [Clostridium perfringens]	74	62	237
2959	1	798	400	[gi 1405454	[aconitase [Bacillus subtilis]	74	60	399
2962	1	650	1363	[gi 450866	[3-phosphoglycerate kinase [Thermotoga maritima]	74	58	2HR
2963	1	3	191	[gi 11301893	[YQH [Bacillus subtilis]	74	56	189
3018	1	2	223	[gi 143040	[glutamate-semialdehyde 2,1-aminotransferase [Bacillus subtilis] [pirf0278 042728 Glutamate-semialdehyde 2,1-aminotransferase (EC 4.3.6) - Bacillus subtilis]	74	56	222
3038	1	510	256	[pirf1591 5229	[nitrate reductase alpha chain - Bacillus subtilis (fragment)	74	57	255
3062	1	374	169	[gi 1107528	[lrg start [Corynebacterium]	74	51	186
4035	1	184	360	[gi 1022725	[Unknown [Streptococcus haemolyticus]	74	64	177
4045	1	607	305	[gi 1510977	[M. jannaschii predicted coding region K10918 [Methanococcus jannaschii]	74	41	303
4283	1	471	304	[gi 520844	[orf4 [Bacillus subtilis]	74	58	168
4449	1	3	221	[gi 580910	[peptide-synthetase ORF1 [Bacillus subtilis]	74	54	219
4587	1	458	231	[gi 1370207	[orf6 [Laccobacillus sakei]	74	59	226

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4603	1	29	214	gi 146208	glutamate synthase large subunit (EC 2.6.1.53) [Escherichia coli] pir A39617 A29617 glutamate synthase [NADPH] (EC 1.4.1.11) large chain - Escherichia coli	74	60	186
4670	1	166	184	gi 1256135	[YbbF] [Bacillus subtilis]	74	61	183
5	110	7953	7162	gi 143727	[putative] [Bacillus subtilis]	73	42	792
11	2	2454	1372	gi 166338	[dihydroorotate dehydrogenase] [Agrobacterieae] [Agrobacterieae]	73	55	1081
14	1	2024	1020	gi 143373	[phosphoribosyl aminonucleotide carboxy formyl ornithotransf erase/inosine monophosphate cyclhydrolase (PUR-Hd)] [Bacillus subtilis]	73	54	1005
23	5	5426	4635	gi 1468339	[meso-2,3-butanediol dehydrogenase (D-acetoin forming)] [Klebsiella pneumoniae]	73	58	792
23	117	17739	16360	gi 1297060	[ornithine cyclodeaminase] [Rhizobium meliloti]	73	37	1020
29	2	692	1273	gi 1467442	[stage V sporulation] [Bacillus subtilis]	73	54	582
31	5	6467	4914	gi 1414000	[ipa-76d gene product] [Bacillus subtilis]	73	55	1554
37	9	8658	7403	gi 1429259	[ipa7 gene product] [Bacillus subtilis]	73	59	1257
37	9	7738	7562	gi 168367	[alpha-1,3propylmalate isomerase (putic); putative] [Rhizomucor iranicelloides]	73	52	177
38	7	3931	4896	gi 1050885	[lysN] [Escherichia coli]	73	58	966
44	6	5041	4228	gi 150895	[unknown] [Bacillus subtilis]	73	53	804
44	11	7767	8306	gi 142009	[anabB gene product] [Escherichia coli]	73	50	540
45	3	2419	3080	gi 1109685	[ProN] [Bacillus subtilis]	73	47	642
54	13	14016	113794	gi 413931	[ipa-7d gene product] [Bacillus subtilis]	73	61	243
59	4	1430	2248	gi 147923	[threonine dehydratase 2 (EC 4.2.1.16)] [Escherichia coli]	73	53	819
65	1	1458	710	gi 167794	[ApPF] [Bacillus subtilis]	73	56	729
66	2	1375	860	gi 1560932	[mrd gene product] [Bacillus subtilis]	73	53	516
102	13	10124	11179	gi 500891	[3-isopropylmalate dehydrogenase (AA 1 - 165) [Bacillus subtilis] - acillus subtilis]	73	55	1056
109	2	3193	12600	gi 1510849	[H. jannaschii predicted coding region HJ0775] [Methanococcus jannaschii]	73	40	894
120	8	4782	5756	gi 1146970	[ribonucleoside triphosphate reductase (Escherichia coli) pir A7331 A47331 anaerobic ribonucleotide reductase - Escherichia coli]	73	56	975
120	9	5726	6223	gi 1204333	[anaerobic ribonucleoside-triphosphate reductase (Haemophilus influenzae)]	73	62	498

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
132	5	4151	4363	[gi 1871048	HPRK2 - heavy chain potential motor protein (Giardia intestinalis)	73	43	213
160	6	5952	4324	[gi 1634107	KcpB (Escherichia coli)	73	59	1629
142	6	7010	5919	[gi 1410125	[trifl gene product (Bacillus subtilis)	73	57	1122
149	4	1866	1717	[gi 1480892	heparin binding protein-44, HBP-44 (Mice, Peptide, 360 aa)	73	53	150
				[gi 10281100801	heparin-binding protein-44 precursor - mouse	73	,	
				[gi 220434	ORF (Mus musculus) (SUS 2-360)			
158	1	1	1431	[gi 162504	ORF-E560 (Escherichia coli)	73	57	1431
174	6	5352	4525	[gi 146240	ketopanturate hydroxymethyltransferase (Bacillus subtilis)	73	55	828
175	8	5537	5178	[gi 1854657	[Na/H antiporter system ORF3 (Bacillus sphaericus)	73	56	360
186	5	6593	5493	[gi 1467477	Unknown [Bacillus subtilis]	73	48	1101
249	6	6283	5729	[gi 1324397	[glycine betaine transporter Opd (Bacillus subtilis)	73	56	555
265	4	1871	2280	[gi 139848	IU3 [Bacillus subtilis]	73	41	408
270	1	328	582	[gi 170461	1220 kDa polyprotein [African swine fever virus]	73	51	255
278	4	4281	3618	[gi 120865	[hypothetical 23.3 kJ protein (Escherichia coli)	73	49	666
279	3	4984	3593	[gi 1185388	[isochorismate synthase (Bacillus subtilis)	73	58	1392
291	4	1207	1575	[gi 1511440	[glutamine--fructose-6-phosphate transaminase (Methanococcus jannaschii)	73	63	369
299	2	735	1166	[gi 147417	Unknown [Bacillus subtilis]	73	51	432
299	5	2050	1324	[gi 467439	[temperature sensitive cell division (Bacillus subtilis)	73	53	1145
334	1	1237	728	[gi 536655	[ORF YB244w (Saccharomyces cerevisiae)	73	43	510
336	2	1827	1036	[gi 70943	[urea amidolyase [Bacillus subtilis]	73	51	792
374	3	1389	1874	[gi 1403451	[YneJ [Bacillus subtilis]	73	55	486
433	4	1916	2554	[gi 473902	[alpha-acetolactate synthase (Lactococcus lactis)	73	54	639
509	2	1795	1102	[gi 467483	Unknown [Bacillus subtilis]	73	56	768
513	1	1709	918	[gi 1146220	[NAD ⁺ dependent glycerol-3-phosphate dehydrogenase (Bacillus subtilis)	73	56	792
533	2	239	733	[gi 1510605	[hypothetical protein (sp P42287) (Methanococcus jannaschii)	73	44	495
546	2	1148	2835	[gi 43748	[hsm protein (AA 1-520) [Escherichia coli]	73	52	1688
549	1	762	382	[gi 1331847	[ClnA [Bacillus subtilis]	73	57	381
567	1	1346	675	[gi 410137	[ORFX13 [Bacillus subtilis]	73	58	672

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)	
716	1	654	1112	gi 1156623	[exodoxynuclease (Bacillus subtilis)]	73	56	459	
772	1	3	677	gi 112010	Shows 70.2% similarity and 48.6% identity to the Env protein of <i>Almonella</i> typhimurium (Anabaena sp.)	73	57	675	
774	1	3	209	gi 1409286	[bmr (Bacillus subtilis)]	73	52	207	
782	1	1	402	gi 143320	[gap1 gene products (Bacillus megaterium)]	73	56	402	
783	2	451	762	gi 103246	low homology to P4 protein of <i>Haemophilus influenzae</i> and 14.2 kDa protein of <i>Escherichia coli</i> (Bacillus subtilis)	73	56	312	
796	1	3	911	gi 153754	[ABC transporter (Bacillus subtilis)]	73	58	909	
806	3	1209	949	gi 111786	[tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Bacillus subtilis) - <i>Bacillus subtilis</i> pir 37081 YWB cryptophan-tRNA ligase (EC 6.1.1.2)]	73	51	261	
816	2	4839	1	3097	gi 141748	[hsdh protein (AA 1-520) (Escherichia coli)]	73	52	1743
839	1	798	400	gi 886906	argininonucleotide synthetase (Streptomyces clavuligerus) pir S57659 IS57659	73	59	399	
857	1	3	290	gi 146052	[acetoin utilization protein (Bacillus subtilis)]	73	50	288	
1008	1	790	398	gi 140100	rodc (tag) polypeptide (AA 1-746) [Bacillus subtilis] pir S06149 IS06019	73	41	393	
1018	1	1	213	gi 1529357	argininoacaine synthase (EC 6.3.4.5) - treponema clavuligerus	73	53	213	
1011	1	3	491	gi 1142706	[comC gene product (Bacillus subtilis)]	73	51	419	
1174	1	395	1	204	gi 1149513	[alpha/beta subunit of laminin 5 (Homo sapiens)]	73	60	192
1175	1	655	1	329	gi 1172817	[No definition line found (Ctenorhynchus elegans) sp P46975 STT3_CAREL	73	57	327
1187	1	3	209	gi 1580870	[ipa-37d gene product (Bacillus subtilis)]	73	52	207	
1206	1	72	245	gi 144816	[formyltetrahydrofolate synthetase (FTHFS) (tgg start codon) (EC 4.4.1)]	73	43	174	
1454	1	423	1	241	gi 11212253	[unknown (Schirosaecharamyces pombe)]	73	53	183
1469	1	517	1	260	gi 1130787	[YqeG (Bacillus subtilis)]	73	55	256
1761	1	374	1	189	gi 91915	[Msr26Aa gene product (Drosophila simulans)]	73	34	186
1849	1	467	1	243	gi 162207	[DNA topoisomerase II (Trypanosoma cruzi)]	73	60	225
2055	1	2	1	400	gi 152281	[P47K protein (Rhadococcus erythrophilus)]	73	34	399
2556	1	2	1	244	gi 145925	[fec8 (Escherichia coli)]	73	62	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (10 Int)	Stop (10 Int)	match accession	match gene name	% sim	% ident	length (nt)
2947	2	549	400	gi 114660	[polynucleotide phosphorylase (Bacillus subtilis)]	71	51	150
2956	1	746	375	gi 143397	[quinol oxidase (Bacillus subtilis)]	73	58	372
3037	1	655	329	gi 1111091	[lactolactate synthase (Bacillus subtilis)]	73	55	327
3115	1	385	194	gi 1323866	[overlapping out-of-phase protein (Eggplant mosaic virus). sp P01291 VTK_EPVK_70 kD PROTEIN.]	73	53	192
3603	2	700	527	gi 1419521	[glutaryl-CoA dehydrogenase precursor (Mus musculus)]	73	48	174
3743	1	798	400	gi 1506688	[highly gene of Ecopcr1 gene product (Escherichia coli) pir S8437 S8437 hadh]	73	54	399
					[protein - Escherichia coli pir S0932 S0932 hypothetical protein A - Escherichia coli (SUB 48-54.0)]			
3752	1	660	359	gi 1524193	[unknown (Mycobacterium tuberculosis)]	73	59	282
3852	1	2	181	gi 216746	[D-lactate dehydrogenase (Lactobacillus plantarum)]	73	68	180
3914	1	475	239	gi 1511450 S134	[Hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - Chicken (fowl)]	73	53	227
3914	2	1570	343	gi 1526991	[unknown (Bacillus subtilis)]	73	38	228
4069	1	2	316	gi 14003	[oxoglutarate dehydrogenase (NADP+) (Bacillus subtilis) gi 23129 0001_BACSU 2-OXOCYCLATE DEHYDROGENASE EI COMPONENT (EC 2.4.2) (ALPHA-KETOGLUTARATE DEHYDROGENASE).]	73	55	315
4165	1	715	365	gi 1419521	[glutaryl-CoA dehydrogenase precursor (Mus musculus)]	73	48	351
4196	1	1	177	gi 109660	[deoxyribose-phosphate aldolase (Bacillus subtilis) pir S4955 S4955 deoxyribose-phosphate aldolase (EC 4.1.2.4) - <i>Escherichia coli</i> 5]	73	60	177
4202	1	572	376	gi 1528891	[unknown (Bacillus subtilis)]	73	34	195
4314	1	2	193	gi 14336797	[N-acyl-L-amino acid aminohydrolase (Bacillus stearothermophilus) sp P21112 AMA_BACST N-ACYL-L-AMINO ACID AMINOHYDROLASE (EC :5.1.14) (AMINOCYLASE).]	73	47	192
4393	1	3	263	gi 216267	[ORF2 (Bacillus megaterium)]	73	47	261
35	2	903	1973	gi 1114196	[phosphoglycerate dehydrogenase (Bacillus subtilis)]	72	53	1071
38	22	19034	17877	gi 1602031	[similar to trimethylamine DH (Mycoplasma capricolum) pir S8950 S8950 probable trimethylamine dehydrogenase (EC :5.99.7) - <i>Mycoplasma capricolum</i> (SCC) (fragment)]	72	54	1218
38	13	18134	19162	gi 1113968	[ipa44d gene product (Bacillus subtilis)]	72	54	1039
44	19	11895	122953	gi 1516272	[unknown (Bacillus subtilis)]	72	49	1059
48	7	6346	7117	gi 11459	[lipoate synthase (Methanobacterium halobium)]	72	49	870
50	7	6563	5691	gi 11205399	[proton glutamate symporter protein (Haemophilus influenzae)]	72	51	873

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
53	9	10321	9259	[gi 1310956	[TqjE [Bacillus subtilis]	72	52	1263
56	123	29349	29995	[gi 1467471	[Unknown [Bacillus subtilis]	72	47	447
69	4	5298	4123	[gi 11351725	[InfoR [Treponema pallidum]	72	46	1176
69	5	4377	4982	[gi 904198	[Hypothetical protein [Bacillus subtilis]	72	43	606
73	1	2	856	[gi 142997	[Glycerol uptake facilitator [Bacillus subtilis]	72	59	855
98	133	9371	10258	[gi 1467335	[Unknown [Bacillus subtilis]	72	50	688
127	1	1	1593	[gi 1217144	[Alanine carrier protein [thermophilic bacterium PS3] piri[A45111][A45111]	72	56	1593
131	1	5197	2600	[gi 153952	[Polymerase III polymerase subunit 1' [Salmonella typhimurium] - Salmonella typhimurium piri[A45111][A45115] DNA-directed RNA polymerase (EC 2.7.7.7) 1' alpha chain	72	53	2398
141	4	1040	1978	[gi 1405446	[Transketolase [Bacillus subtilis]	72	54	919
149	8	2819	2535	[gi 606334	[Isocy [Escherichia coli]	72	44	285
149	177	5472	5265	[gi 1304472	[DNA polymerase I unidentified phycodnavirus clone ORF41	72	55	228
154	1	1	210	[gi 1205620	[Ikeritin like protein [Haemophilus influenzae]	72	40	210
155	1	2207	1120	[gi 1391610	[Farnesyldiphosphate synthase [Bacillus stearothermophilus] piri[320571][32057] geranyltransferase (EC 2.5.1.10) - bacillus stearothermophilus	72	57	688
180	1	2	328	[gi 1313630	[A180 [Saccharomyces cerevisiae]	72	62	327
184	3	1145	3553	[gi 1203110	[Virulence associated protein homolog [Haemophilus influenzae]	72	49	2109
195	2	1923	1279	[gi 101730	[Hypothetical protein [Syncytiotrophoblast cell line]]	72	45	645
206	113	14646	18589	[gi 1066807	[Ornithine amidotransferase [Bacillus subtilis]	72	50	1224
209	2	462	932	[gi 1204666	[Hypothetical protein (GB:K73124-5)] [Haemophilus influenzae]	72	60	471
215	2	764	522	[gi 881513	[Insulin receptor homolog [Drosophila melanogaster] piri[57245][557245] insulin receptor homolog - fruit fly [Drosophila melanogaster] (SUB 46-2146)	72	63	243
224	1	2	790	[gi 949974	[Sucrose repressor [Staphylococcus xylosus]	72	54	789
233	1	1526	765	[gi 1408493	[Homologous to SwissProt YIDA_ECOLI hypothetical protein [Bacillus subtilis] (ORF_o470) [Escherichia coli]	72	52	762
240	1	220	1485	[gi 537049	[Hypothetical protein (GB:006349_1) [Haemophilus influenzae]	72	52	1266
245	1	3	1340	[gi 120578		72	46	1338

TABLE 2

5
10
15
20
25
30
35
40
45
50
55

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start	Stop	match	match	% sim	% ident	length
In	ID	Intr	Intr	accession	accession			(nt)
259	2	2108	1245	[gi 1340128]	[ORF1 (Staphylococcus aureus)]	72	59	864
104	2	265	1094	[gi 1205130]	[leucine-binding periplasmic protein (Haemophilus influenzae)]	72	52	810
307	130	5326	5039	[gi 1070015]	[protein-dependent (Bacillus subtilis)]	72	53	286
315	1	517	260	[gi 143399]	[quinol oxidase (Bacillus subtilis)]	72	55	258
316	13	932	9308	[gi 1204445]	[hypothetical protein (S. P277857) (Haemophilus influenzae)]	72	58	315
337	3	926	1609	[gi 1487433]	[citrate synthase II (Bacillus subtilis)]	72	55	684
364	7	12538	10193	[gi 1510643]	[ferric iron transport protein B (Methanococcus Jannaschii)]	72	53	206
409	2	340	1263	[gi 1402944]	[orfM1 gene product (Bacillus subtilis)]	72	49	324
441	3	2177	1520	[gi 1312379]	[highly conserved among subacteria (Clostridium acetobutylicum)]	72	48	581
453	6	2654	2505	[pir S000601 BMSA]	[antibacterial protein 3 - Staphylococcus haemolyticus]	72	70	150
460	1	2	625	[gi 1016162]	[ABC transporter subunit (Cyanophora paradoxa)]	72	51	624
463	1	3253	1628	[gi 666014]	[The polymorphism (RFIP) of this gene is associated with acceptability to CoA synthetase (Homo sapiens)]	72	60	1626
480	4	3047	1466	[gi 1433992]	[ATP synthase subunit epsilon (Bacillus subtilis)]	72	51	420
502	1	1086	586	[gi 130059]	[ORF2 (Synchococcus sp.)]	72	50	501
519	1	81	1184	[gi 1303704]	[Yke (Bacillus subtilis)]	72	54	1104
559	1	3	746	[gi 1107530]	[leuD gene product (Corynebacterium)]	72	56	744
575	1	1142	573	[gi 1303866]	[fogs (Bacillus subtilis)]	72	56	570
671	1	2	592	[gi 1204497]	[protein-in-export membrane protein (Haemophilus influenzae)]	72	44	591
679	2	295	1251	[gi 153358]	[virulence-associated protein E (Dichelobacter nodosus)]	72	52	957
687	2	295	957	[gi 1146214]	[48% identical amino acids with the Escherichia coli smb suppressor; putative (Bacillus subtilis)]	72	49	663
837	1	1	435	[gi 1146183]	[putative (Bacillus subtilis)]	72	54	415
868	1	150	788	[gi 1377842]	[unknown (Bacillus subtilis)]	72	55	619
922	1	130	432	[gi 1088269]	[unknown protein (Acinetobacter vinelandii)]	72	58	303
941	1	2	238	[gi 1533299]	[NADPH-pulifite reductase flavoprotein component (Salmonella typhimurium)]	72	49	237
980	1	840	421	[gi 853767]	[UDP-N-acetylglucosamine 1-carboxyvinyltransferase (Bacillus subtilis)]	72	59	420

TABLE 2

5. *s. aureus* - Putative coding regions of novel proteins similar to known proteins5
10
15
20
25
30
35
40
45
50

Contig	ORF ID	Start ID (nt)	Stop ID (nt)	match accession	match gene name	% sim	% ident	length (nt)
1209	1	383	213	[gi 1144735]	neurotoxin type B [Clostridium botulinum]	72	44	171
1469	1	671	474	[gi 1105458]	hypothetical protein (GB:D2652-17) [Haemophilus influenzae]	72	63	198
1956	1	727	365	[gi 1134409]	hexosephosphate transport protein [Salmonella typhimurium] DipB4851 [Sal83 hexose phosphate transport system regulatory protein uhpB]	72	44	363
2101	1	3	401	[gi 11303950]	[Yqly] [Bacillus subtilis]	72	50	399
2503	1	569	199	[gi 119713]	formate dehydrogenase [Methanobacterium formicum] DicM2712 [A42712] formate dehydrogenase [EC 1.2.1.2] - ethanobacterium formicum	72	56	171
2867	1	3	155	[gi 11212729]	[Yqbj] [Bacillus subtilis]	72	46	153
3004	1	167	185	[gi 1659399]	hypothetical protein [Bacillus subtilis]	72	55	183
3109	1	278	141	[gi 1413968]	[Ipa-44d gene product] [Bacillus subtilis]	72	45	138
3171	1	3	287	[gi 1515938]	glutamate synthase (ferredoxin) [Synechocystis sp. P-1] pirS16957 [S46957] glutamate synthase (ferredoxin) [EC 1.4.7.1] - cyanobacteria sp.	72	52	285
3771	1	26	167	[gi 1408501]	homologous to N-acyl-L-amino acid amidohydrolase of <i>Bacillus stearothermophilus</i> [Bacillus subtilis]	72	63	342
3951	1	1	222	[gi 1500409]	Iw_jannaschi predicted coding region MJ1519 [Methanococcus jannaschii]	72	38	222
4190	1	721	362	[gi 139956]	[IigC] [Bacillus subtilis]	72	57	360
4444	1	3	247	[gi 11009366]	Respiratory nitrate reductase [Bacillus subtilis]	72	55	345
6	2	911	1200	[gi 157095]	[methylthio carbamoyl transferase] [Bacillus subtilis]	71	46	270
11	115	11350	10859	[gi 1532309]	125 kDa protein [Escherichia coli]	71	47	492
19	1	1248	1235	[gi 11244574]	D-alanine-D-alanine ligase [Enterococcus hitai]	71	52	1188
21	2	898	1488	[gi 196339]	anthranilate synthase component 2 [Leptospira biflexa] pirC2801C32640	71	45	591
34	1	1	567	[gi 11303983]	anthranilate synthase component II [Leptospira biflexa]	71	59	567
37	3	3192	2806	[gi 109681]	[Yqhf] [Bacillus subtilis]	71	50	387
38	18	112250	12462	[gi 1977645]	[arginyl endopeptidase] [Porphyromonas gingivalis]	71	50	213
39	3	1246	4431	[pir S0941 S094]	[spolliE protein - Bacillus subtilis]	71	49	3186
53	14	115770	14760	[gi 1122611]	branched chain alpha-keto acid dehydrogenase E1-alpha [Bacillus subtilis]	71	58	1011
54	11	113461	112625	[pir 113014]	[ont repressor] [Bacillus subtilis]	71	46	837

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start Int.	Stop Int.	match accession	match gene name	% sim	% ident	length (nt)
57	7	7152	5860	gi 1503175	EFIC domain of PTS-dependent Gat transport and phosphorylation Escherichia coli	71	48	1231
57	18	11897	11334	gi 1033247	high homology to flavohemoprotein (HemeoGlobin-like protein) of <i>Alcaligenes</i>	71	56	418
62	16	9831	10955	gi 1303946	high homology to flavohemoprotein (HemeoGlobin-like protein) of <i>Alcaligenes</i>	71	54	1125
70	12	8505	8966	gi 147198	[Yqo] (Bacillus subtilis)	71	38	71
86	5	2394	2089	gi 190205	[hypothetical protein] (Bacillus subtilis)	71	51	306
96	7	7601	8269	gi 70991	[hypothetical protein] (Bacillus subtilis)	71	49	669
100	6	4822	5931	gi 100848	[hypothetical protein] (Bacillus subtilis)	71	45	1110
103	1	1062	532	gi 14089	[hypothetical protein] (Bacillus subtilis)	71	41	531
109	18	15312	15695	gi 141985	[ipa-fid gene product] (Bacillus subtilis)	71	57	384
113	1	630	316	gi 66254	[probable protein kinase] (Saccharomyces cerevisiae)	71	57	315
114	5	6598	5603	gi 143156	[membrane bound protein] (Bacillus subtilis)	71	40	936
113	2	3087	1723	gi 130913	[QhX] (Bacillus subtilis)	71	53	1365
149	19	6315	5095	gi 529650	[C40P] (Bacteriophage SP1)	71	51	441
154	5	3635	3087	gi 425468	[repressor protein] (Streptococcus sobrinus)	71	47	549
164	11	11354	11689	gi 49318	[OKF4 gene product] (Bacillus subtilis)	71	52	116
169	5	1936	2745	gi 1401403	[Unknown] (Mycobacterium tuberculosis)	71	56	810
193	2	272	1234	gi 1303788	[Yqeh] (Bacillus subtilis)	71	49	963
205	1	1743	895	gi 125694	[GlnQ] (Mycoplasma pneumoniae)	71	46	849
233	4	1849	2022	gi 633732	[ORF1] (Candybacter jejuni)	71	50	174
237	7	4501	5169	gi 1449384	[HistE] (Lactococcus lactis)	71	54	669
272	4	2848	2273	gi 170993	[hypothetical protein] (Bacillus subtilis)	71	48	576
274	2	618	1496	gi 143035	[NAD(P)H:glutamyl-transfer RNA reductase] (Bacillus subtilis)	71	53	879
276	5	3349	2720	gi 130362	[ORF210] (Escherichia coli)	71	50	630
287	1	116	660	gi 110634	[pirA522] (A522 5'-aminolevulinate synthase) - ecilus	71	53	525
288	6	3322	2771	gi 1256625	[leutative] (Bacillus subtilis)	71	47	552

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
301	6	3432	2461	[gi 467417	similar to lysine decarboxylase [Bacillus subtilis]	71	57	1032
306	4	6607	5222	[gi 1556618	transport protein [Bacillus subtilis]	71	56	1186
307	2	1536	925	[gi 632633	[orfC [Mycoplasma capricolum]	71	45	612
310	5	5793	5146	[gi 348052	lacton utilization protein [Bacillus subtilis]	71	51	648
322	1	2	1303	[gi 1101819	hypothetical protein [Synechocystis sp.]	71	46	1102
333	4	4171	1995	[gi 467473	unknown [Bacillus subtilis]	71	57	177
350	2	546	922	[gi 551879	[orfF 1 [Lactococcus lactis]	71	55	375
375	4	1860	3071	[gi 467447	unknown [Bacillus subtilis]	71	57	1212
380	5	1560	2102	[gi 125557	[AMP synthase b subunit [Bacillus megaterium]	71	43	543
414	2	251	637	[gi 50904	homologous to E. coli rmpA [Bacillus subtilis]	71	49	387
424	1	335	1354	[gi 551305	L-lactate dehydrogenase [Lactobacillus plantarum]	71	57	1020
436	4	3701	3270	[pir PK0501 PN05	phosphoribosylanthranilate isomerase (EC 5.3.1.24) - Bacillus subtilis	71	66	432
482	1	3	1280	[gi 410142	[orfX18 [Bacillus subtilis]	71	49	1778
525	3	2272	1844	[gi 1133370	phosphoribosylpyrophosphate amidotransferase (PRN-F; EC 2.4.2.14) Bacillus subtilis	71	56	429
529	4	2739	2047	[gi 606150	[orf_F109 [Escherichia coli]	71	41	691
563	1	22	969	[gi 12337015	[orfF4 [Bacillus subtilis]	71	53	948
581	1	506	255	[gi 1101710	[orf25G1.2 [Candidorhabditis elegans]	71	47	252
612	2	1068	913	[gi 155968	fimbriae Z [Salmonella typhimurium]	71	55	156
613	1	1	654	[gi 466778	lysine specific permease [Escherichia coli]	71	50	654
618	1	1243	623	[gi 1146238	[polyIA] polymerase [Bacillus subtilis]	71	52	621
620	1	1170	586	[gi 1086243	unknown [Bacillus subtilis]	71	53	585
691	1	1126	641	[gi 209260	core ORF1 [Bacillus subtilis]	71	51	486
694	2	149	427	[gi 12971	NADH dehydrogenase subunit V (AA 1-605) [Gallus gallus] ir S1097 S10197	71	47	279
					NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain - chicken mitochondrial (S021)			
715	2	169	777	[gi 10330	YofL [Bacillus subtilis]	71	53	609
746	2	1473	970	[gi 137843	unknown [Bacillus subtilis]	71	52	504

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
748	1	1437	802	gi 105659	Yns [Bacillus subtilis]	71	49	636
753	1	1018	524	gi 151089	[H. jannaschii] predicted coding region N00296 [Methanococcus jannaschii]	71	53	495
761	1	3	215	gi 1475912	[pentafunctional enzyme [Pneumocystis carinii]	71	47	213
783	1	1203	703	gi 536635	[ORF YBR244w [Saccharomyces cerevisiae]	71	52	301
800	3	1292	987	gi 1204226	[tRNA delta(2)-isopentenylpyrophosphate transferase [Haemophilus influenzae]	71	48	306
806	1	116	286	gi 141975	[cbhII gene product [Methanobacterium thermophilicum]	71	50	171
931	1	973	488	gi 893388	[Pga [Bacillus subtilis]	71	56	486
1041	1	2	262	gi 1408507	[pyrimidine nucleoside transport protein [Bacillus subtilis]	71	45	261
1070	1	2	172	gi 709933	[hypothetical protein [Bacillus subtilis]	71	46	171
1176	1	57	165	gi 1512559	[HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas movalonii] PtxA4756 A44756	71	49	309
1181	1	366	184	gi 146971	[hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	71	50	161
1261	1	3	290	gi 1153016	[ORF 419 protein [Staphylococcus aureus]	71	50	288
1348	1	456	229	gi 1602681	[orfC [Hycoplana capitulum]	71	48	228
2002	1	756	379	gi 1008177	[ORF YOL046w [Saccharomyces cerevisiae]	71	48	318
2119	1	2	217	gi 1104688	[arginyl-tRNA synthetase [Hycoplana capitulum]	71	50	216
2418	1	3	320	gi 1499771	[H. jannaschii] predicted coding region N10936 [Methanococcus jannaschii]	71	57	318
2961	1	2	187	gi 1312443	[carbamoyl-phosphate synthase (glutamine-hydrolyzing) [Bacillus alidolyticus]	71	57	186
2999	1	2	67	gi 1710020	[nitrite reductase (nirB) [Bacillus subtilis]	71	43	240
3033	1	2	184	gi 11262335	[YmaA [Bacillus subtilis]	71	57	183
3584	1	3	138	gi 1401716	[beta-isopropylmalate dehydrogenase [Neurospora crassa]	71	55	336
3715	1	2	743	gi 1563952	[gluconate permease [Bacillus licheniformis]	71	59	345
3785	1	770	187	gi 147382	[acyl-CoA-dehydrogenase [Streptomyces purpurascens]	71	57	384
3875	1	541	272	gi 11001541	[hypothetical protein [Synchocystis sp.]	71	38	270
4135	1	637	320	gi 142695	[S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase Bacillus megaterium]	71	52	318
4249	1	63	239	gi 1205363	[deoxyribose aldolase [Haemophilus influenzae]	71	63	177
4508	1	530	267	gi 11397667	[vitellogenin [Anolis pulchellus]	71	46	264

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
6	1 3	1 1237	1 2721	[gi 1321788	arginine ornithine antipporter [Clostridium perfringens]	70	54	1485
11	1 13	1 6572	1 7486	[gi 121654	[P]7K [Pseudomonas chlororaphis]	70	41	915
12	1 1	1 2890	1 481	[gi 1467330	replicative DNA helicase [Bacillus subtilis]	70	49	1410
15	1 1	1 1756	1 893	[gi 151216	mannosephosphate isomerase [Streptococcus mutans]	70	46	864
15	1 2	1 1277	1 1050	[gi 476092	Unknown [Bacillus subtilis]	70	50	228
17	1 2	1 2132	1 1350	[gi 145402	choline dehydrogenase [Escherichia coli]	70	52	783
21	1 1	1 2	1 925	[gi 149316	[anthranilate synthase alpha subunit [Lactococcus lactis] pir S35124 S55124	70	50	924
					[anthranilate synthase (EC 4.1.3.27) alpha chain - actococcus lactic subsp. lactic]			
25	1 7	1 5580	1 6251	[gi 1385459	[orfF3 [Bacillus subtilis]	70	52	672
33	1 6	1 6071	1 7423	[gi 1303675	[YgbB [Bacillus subtilis]	70	51	1353
36	1 2	1 1959	1 1594	[gi 500755	[methyl purine glycosylase [Hus musculus]	70	47	636
38	1 6	1 4901	1 5860	[gi 140507	[purimidine nucleoside transport protein [Bacillus subtilis]	70	44	260
44	1 8	1 5312	1 599	[gi 1006620	[hypothetical protein [Synechocystis sp.]	70	49	678
46	1 10	1 8950	1 10030	[gi 1403126	[ecCD gene product [Alcaligenes eutrophus]	70	45	1071
52	1 2	1 2727	1 1900	[gi 1468247	Unknown [Bacillus subtilis]	70	53	828
52	1 6	1 4948	1 4656	[gi 1244501	[esterase II-carboxylesterase (EC 3.1.1.1) [Pseudomonas fluorescens, sp.], [218 aa]	70	50	609
56	1 8	1 8460	1 9962	[gi 1133951	[small subunit of NADH-dependent glutamate synthase [Pleotrichia boryanum]	70	51	1563
62	1 1	1 48	1 290	[gi 1142702	[A competence protein 2 [Bacillus subtilis]	70	47	243
64	1 1	1 1080	1 541	[gi 1204377	[molybdopterin biosynthesis protein [Haemophilus influenzae]	70	47	540
70	1 5	1 5139	1 3555	[gi 1204834	[2',3'-cyclic-nucleotide 2'-phosphodiesterase [Haemophilus influenzae]	70	47	1545
91	1 4	1 7793	1 5466	[gi 886471	[methionine synthase [Catharanthus roseus]	70	56	2328
96	1 5	1 8754	1 7255	[pir B3996 B390	[alkaline phosphatase (EC 3.1.3.1) III precursor - Bacillus subtilis	70	54	1500
110	1 2	1 767	1 1300	[gi 115294	[adenine phosphoribosyl-transferase [Escherichia coli]	70	51	534
116	1 6	1 7026	1 7976	[gi 143667	[poruloprotein [Bacillus subtilis]	70	50	951
121	1 8	1 6401	1 6988	[gi 1107228	[tcp start [Caulobacter coli]	70	45	588
131	1 8	1 6842	1 7936	[gi 1150454	[prolidase PepO [lactobacillus delbrueckii]	70	48	1095

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
135	1	2	1489	[gi 3111309	[putative membrane-bound protein with four times repetition of ro-Ser-Ala at the N-terminus; function unknown [Alcaligenes utopius]	70	49	1488
138	1	418	714	[gi 190181	[hypothetical protein [Bacillus subtilis]	70	46	297
164	8	9344	9874	[gi 4915	[ORF1 gene product [Bacillus subtilis]	70	47	531
164	16	11526	11618	[gi 1205212	[hypothetical protein [GB.D10483_181] [Haemophilus influenzae]	70	50	993
205	2	2735	1803	[gi 1215695	[peptide transport system protein sapf homolog; Sapf homolog [Mycoplasma pneumoniae]	70	47	933
209	3	910	1186	[gi 1204665	[hypothetical protein [GB.X7124_26] [Haemophilus influenzae]	70	48	477
246	3	1340	756	[gi 1215098	[excisionase [Bacteriophage 15a]	70	46	417
263	7	7876	6749	[gi 1147540	[Aspartokinase II [Bacillus sp.]	70	51	1128
268	3	3212	4117	[gi 1340128	[ORF1 [Staphylococcus aureus]	70	50	906
302	1	6	3201	[gi 147782	[trna protein (trna start) [Escherichia coli]	70	46	627
302	10	5879	7051	[gi C18510C45	[tRNA-r1 (bosy)transferase (EC 2.4.2.29) - Escherichia coli	70	55	1173
313	1	2520	1414	[gi 1205934	[aminopeptidase a1 [Haemophilus influenzae]	70	46	1107
355	2	379	659	[gi 1070013	[protein-dependent [Bacillus subtilis]	70	48	291
403	1	1255	629	[gi 7331147	[GufF (Xanthomonas campestris)	70	33	627
444	10	8770	7273	[gi 1204752	[high affinity ribose transport protein [Haemophilus influenzae]	70	52	504
449	1	2	1243	[gi 1613724	[Hfq [Bacillus firmus]	70	44	1244
472	1	637	120	[gi 727145	[open reading frame; putative [Bacillus myoliquefaciens] pir 229091 B29091	70	41	318
480	2	727	1608	[gi 1142560	[hypothetical protein (logA region) - Bacillus myoliquefaciens (fragment)]	70	44	882
524	1	2	307	[gi 602292	[RCH2 protein [Brassica napus]	70	45	306
525	1	823	413	[gi 1141372	[phosphotransferase (PUT-NI) [Bacillus subtilis]	70	52	411
565	4	3625	12552	[gi 1881434	[ORF [Bacillus subtilis]	70	51	1074
607	4	629	1284	[gi 311324	[hypothetical protein (SP_P17002) [Methanococcus jannaschii]	70	50	456
633	1	1183	703	[gi 431231	[uracil permease [Bacillus coidolyticus]	70	53	681
646	3	1683	1309	[gi 467340	[unknown [Bacillus subtilis]	70	49	375
663	1	830	417	[gi 1103673	[W92 [Bacillus subtilis]	70	40	414

TABLE 2

Contig	ORF	Start (In Int)	Stop (In Int)	match accession	match gene name	% sim	% ident	length (nt)
661	1	1468	781	[gi 1001678]	[hypothetical protein (Synechocystis sp.)]	70	53	708
708	1	2	448	[sp P13940 YOH_]	[hypothetical 54.3 kD protein in ECO-ALB INTERGENIC REGION]	70	51	447
725	1	51	722	[gi 1001644]	[hypothetical protein (Synechocystis sp.)]	70	48	672
776	1	1371	797	[gi 1153165]	[putative (Escherichia coli)]	70	47	585
814	1	250	783	[gi 1532971]	[NADH dehydrogenase (ndhF) (Vicia faba)]	70	47	534
865	1	1585	1379	[gi 1204636]	[ATP-dependent helicase (Hansenophilus influenzae)]	70	45	207
894	1	535	269	[gi 1457364]	[DNA binding protein (probable) (Bacillus subtilis)]	70	41	267
919	1	3	317	[gi 134847]	[CNA (Bacillus subtilis)]	70	40	315
944	1	3	572	[gi 709991]	[hypothetical protein (Bacillus subtilis)]	70	44	570
988	2	772	605	[gi 142441]	[ORF 3; putative (Bacillus subtilis)]	70	50	168
1055	1	3	335	[gi 1529755]	[spc (Streptococcus pyogenes)]	70	37	133
1093	1	2	904	[gi 1853754]	[ABC transporter (Bacillus subtilis)]	70	49	903
1109	1	2	310	[gi 1001827]	[hypothetical protein (Synechocystis sp.)]	70	42	309
1220	1	468	235	[pir 52316 S234]	[epib protein - Staphylococcus epidermidis]	70	40	234
1279	1	73	348	[gi 153015]	[F400 protein (Staphylococcus aureus)]	70	47	276
1316	1	195	542	[sp P01776 PUPA_]	[PENICILLIN-BINDING PROTEIN 1A (PBP-1A) (PENICILLIN-BINDING PROTEIN A)]	70	50	348
1537	1	232	402	[gi 1146181]	[putative (Bacillus subtilis)]	70	50	171
1574	1	451	272	[gi 219630]	[endothelin-A receptor (Homo sapiens)]	70	47	180
1640	1	690	346	[gi 1146243]	[22.4A identity with Escherichia coli DNA-damage inducible protein ...; putative (Bacillus subtilis)]	70	46	345
2504	1	2	286	[gi 495179]	[transmembrane protein (Lactococcus lactis)]	70	51	285
3061	1	564	101	[gi 508175]	[ELIC domain of PMS-dependent DNA transport and phosphorylation Escherichia coli]	70	44	264
3128	1	2	199	[gi 1340096]	[Unknown (Mycobacterium tuberculosis)]	70	51	198
3218	1	3	488	[gi 1535938]	[glutamate synthase (ferredoxin) (Synechocystis sp.) pir 546937 346557 glutamate synthase (ferredoxin) (EC 1.4.7.1) - <i>ynecho cystis</i> sp.]	70	50	486
3323	1	794	399	[gi 1154691]	[ATP binding protein (Phormidium lesinum)]	70	52	396
3679	1	599	399	[gi 1539365]	[chromosome condensation protein (Caenorhabditis elegans)]	70	40	201

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1841	1	706	198	gi 108965	hypothetical 23.3 kd protein [Escherichia coli]	70	47	309
3929	1	3	401	gi 148435	putative [Lactococcus lactis]	70	49	359
4044	1	595	374	gi 602031	similar to triethylamine Dio [Mycoplasma capricolum] pir 34950 is4950	70	40	222
4339	1	558	280	gi 133951	probable triethylamine dehydrogenase (EC. 5.99.7) - Mycoplasma capricolum (SGC) (fragment)	70	49	279
4422	1	576	289	gi 126464	small subunit of NADH dependent glutamate synthase [Plectonema boryanum]	70	49	288
4667	1	161	200	gi 166412	NADH-glutamate synthase [Medicago sativa]	70	59	162
16	8	7571	9031	gi 1498620	M. jannaschii predicted coding region MJ0798 (Methanococcus jannaschii)	69	44	1461
16	9	9080	10033	gi 1133197	thioredoxin reductase [Escherichia coli/nphilum]	69	54	954
30	1	1452	727	gi 11204910	hypothetical protein [GBU14001 02] [Haemophilus influenzae]	69	52	726
38	4	1023	1298	gi 1407773	[dava gene product] [Anabaena sp.]	69	41	276
44	9	5987	6595	gi 1205920	molybdate uptake system hydrophilic membrane-bound protein [Haemophilus influenzae]	69	45	609
62	15	9104	9475	gi 1863178	Unknown [Bacillus subtilis]	69	44	372
66	4	2402	12803	gi 11301893	[Yqlh] [Bacillus subtilis]	69	51	402
67	15	14124	133027	gi 149647	ORF2 [Listeria monocytogenes]	69	37	494
67	17	14053	14382	gi 105002	[ORF_1356] [Escherichia coli]	69	49	310
67	19	15130	15807	gi 11109684	[Prov] [Bacillus subtilis]	69	45	678
78	3	1447	2124	gi 1256633	[putative] [Bacillus subtilis]	69	53	678
78	4	4513	3725	gi 101958	YQIG [Bacillus subtilis]	69	32	789
85	4	4521	4213	pir E29336 2293	hypothetical protein [pir operon] - <i>Bacillus subtilis</i>	69	32	309
86	6	3253	2654	gi 173332	[OrfC] [Bacillus subtilis]	69	50	600
95	1	96	710	gi 76666	(A) antigen, sperm tail membrane antigen/putative sucrose-specific phosphotransferase enzyme II homolog [rice, testis, Peptide Partial, 72 aa]	69	43	615
100	7	6023	7426	gi 1203555	[Na+/H ⁺] antiporter [Haemophilus influenzae]	69	39	1404
102	2	2678	1450	gi 561690	leucoglycoprotease [Pasteurella haemolytica]	69	47	1029
103	8	12241	18537	gi 1109366	[respiratory nitrate reductase [Bacillus subtilis]]	69	54	3705
103	11	14987	12552	gi 170020	[nitrite reductase [nirB] [Bacillus subtilis]]	69	51	2436

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nc)	Stop (nc)	match accession	match gene name	% sim	% ident	length (nt)
112	11	8708	10168	gi 154111	hexophosphate transport protein (<i>Salmonella typhimurium</i>)	69	51	1461
				gi 101833	hexose phosphate transport system protein upf -			
112	16	16444	17244	gi 11204435	pyruvate formate-lyase activating enzyme (<i>Haemophilus influenzae</i>)	69	50	771
113	2	33	953	gi 1290509	oJ07 (<i>Escherichia coli</i>)	69	43	921
114	2	1537	1058	gi 1A2771 A427	reticulocytobinding protein 1 - <i>Plasmodium vivax</i>	69	39	480
121	6	4309	5310	gi 1154633	[ORF] (<i>Bacillus subtilis</i>)	69	53	1002
125	2	267	1854	gi 141931	lps-7d gene product (<i>Bacillus subtilis</i>)	69	43	568
149	27	10666	11000	gi 1528089 S280	hypothetical protein A - yeast (<i>Zygosaccharomyces bisporus</i>)	69	39	267
161	1	1598	411	gi 1120538	hypothetical protein (GBU14001_3021) (<i>Haemophilus influenzae</i>)	69	47	786
165	4	2222	4633	gi 140054	phenylalanyl-tRNA synthetase beta subunit (AA 1-864) (<i>Bacillus brevis</i>)	69	52	2412
169	3	1210	1761	gi 1296031	elongation factor Te (<i>Spizellina plateaeid</i>)	69	45	552
175	12	8886	8339	gi 1732682	[PME protein] (<i>Escherichia coli</i>)	69	69	348
190	2	484	1671	gi 17731 H158 -	[HISTIDYL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9) (IMIDAZOLE ACETOL-PHOSPHATE TRANSMINASE)]	69	48	1188
206	1	5551	2777	gi 141750	hadr protein (AA 1-1031) (<i>Escherichia coli</i>)	69	49	2775
206	4	6038	5796	gi 1256115	[ybfB] (<i>Bacillus subtilis</i>)	69	48	243
249	1	616	319	gi 1404346	YncP (<i>Yersinia enterocolitica</i>)	69	40	111
302	6	4820	5776	gi 1001768	hypothetical protein (<i>Synechocystis sp.</i>)	69	48	957
324	12	7384	3693	gi 1256798	pyruvate carboxylase (<i>Rhizobium etli</i>)	69	53	3692
351	3	2098	1808	gi 1491664	[TOH1.4] (<i>Cetorhabditis elegans</i>)	69	30	291
369	3	2075	2105	gi 136458	[ORF] (<i>Balenoptera acutorostrata</i>)	69	61	231
392	3	1999	2424	gi 156015	[ORF] (<i>Bacillus subtilis</i>)	69	45	426
410	1	87	779	gi 125611	[phosphoglyceromutase (<i>Zymomonas mobilis</i>)]	69	58	693
421	1	2085	1129	gi 1276985	[larginase (<i>Bacillus caldovelox</i>)]	69	54	957
444	8	6713	7741	gi 1221782	purine synthesis repressor (<i>Haemophilus influenzae</i>)	69	40	1029
453	1	828	415	gi 1122758	unknown (<i>Bacillus subtilis</i>)	69	57	414
469	2	2886	2246	gi 1458228	[muy homolog (<i>Homo sapiens</i>)]	69	44	1041

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	Start (nt)	Stop (nt)	match accession	match gene name	Start (nt)	Stop (nt)	match accession	match gene name
509	3	1730	1371	[gi 19224]	[ORF 4 [Synchococcus sp. 1]							69	19
520	5	3021	2823	[gi 226627]	[similar to D. melanogaster Hsp101-2 protein (PDB:1S1A)]							69	39
531	1	26	760	[gi 509672]	[repressor protein (bacteriophage Tuc009)]							69	33
589	1	107	253	[gi 169101]	[17.9 kDa heat shock protein (hsp17.9) [Pisum sativum]]							69	52
594	2	597	1391	[gi 142783]	[DNA phosphatase (Bacillus firmus)]							69	46
604	4	2476	2114	[gi 113930]	[lipA-6d gene product (Bacillus subtilis)]							69	45
607	1	2	313	[gi 1236103]	[W0802_3 [Caenorhabditis elegans]]							69	47
607	2	590	1312	[gi 516715]	[ORF YBR275c [Saccharomyces cerevisiae]]							69	39
734	1	864	433	[gi 167327]	[unknown (Bacillus subtilis)]							69	44
759	1	3	338	[gi 1003367]	[Respiratory nitrate reductase (Bacillus subtilis)]							69	50
761	2	392	586	[gi 15508]	[Leucyl-tRNA synthetase (cytoplasmic) [Saccharomyces cerevisiae]]	[1370340]						69	336
802	1	72	1013	[gi 143044]	[Ferrochelatase (Bacillus subtilis)]							69	55
816	1	2573	1368	[gi 1510268]	[restriction modification system S subunit (Methanococcus jannaschii)]							69	45
838	2	133	387	[gi 1255371]	[coded for by C. elegans cDNA YK449.5; coded for by C. elegans cDNA YK348.1; Similar to quinolinate kinase [Caenorhabditis elegans]]							69	255
851	2	745	1005	[gi 184898]	[sech gene product (Antennariun sp.)]							69	39
867	1	535	269	[gi 1070014]	[protein-dependent (Bacillus subtilis)]							69	47
995	1	954	478	[gi 1205569]	[transcription elongation factor (Haemophilus influenzae)]							69	53
999	1	1009	506	[gi 189254]	[predicted trithorax protein (Drosophila virilis)]							69	477
1127	1	1315	659	[gi 1205434]	[H. influenzae predicted coding region H1191 (Haemophilus influenzae)]							69	21
1138	1	248	460	[gi 1510646]	[H. jannaschii predicted coding region H40568 (Methanococcus jannaschii)]							69	504
2928	1	3	401	[gi 120503]	[glutamate permease (Escherichia coli)]							69	41
3090	1	444	223	[gi 1204987]	[DNA polymerase III, alpha chain (Haemophilus influenzae)]							69	36
3817	1	2	400	[gi 1461199]	[leucyl-tRNA synthetase (Mycobacteriopsis mediterranei)]							69	222
3933	1	667	335	[gi 1524193]	[unknown (Mycobacterium tuberculosis)]							69	46

TABLE 2

5 10 15 20 25 30 35 40 45 50

55

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4079	1	747	400	gi 546918	orf3' of comX [Bacillus subtilis, E26, Peptide Partial, 140 aa]	69	64	348
					gi 513612 S41612 hypothetical protein Y - <i>Bacillus subtilis</i>			
					sp P03961 YKXD_BACSU HYPOTHETICAL PROTEIN IN COMX' REGION (ORF3')			
					FRAGMENT			
4115	2	215	400	gi 527205	[67.1 kDa Mycosin-crossreactive streptococcal antigen [Streptococcus yoghensi]]	69	59	186
4139	1	1	333	gi 1120451	[hypothetical protein [Synechocystis sp. P-1]	69	36	333
4256	1	457	230	gi 496158	[restriction-modification enzyme subunit M1 [Mycoplasma pneumoniae]]	69	41	224
					gi 569395 S40395 fimbrial protein - <i>Mycoplasma pneumoniae</i> (SPC3)			
4317	1	90	374	gi 4133967	[16-kDa-41-kDa gene product [Bacillus subtilis]]	69	44	285
4465	1	1	293	gi 396296	[similar to phosphotransferase system enzyme II, [Escherichia coli]]	69	49	291
					sp P12672 PMC_ECOLI PTS SYSTEM, FRUCTOSE-LIKE-2 IIC COMPONENT			
					PHOSPHOTRANSFERASE ENZYME II, C COMPONENT			
3	1	2302	1193	gi 1109685	[ProW [Bacillus subtilis]]	68	46	1110
15	4	2592	2074	gi 607973	[Unknown [Saccorhomyces cerevisiae]]	68	45	519
31	8	6328	8772	gi 290642	[AlPase [Enterococcus hirai]]	68	48	2445
40	2	1115	750	gi 606342	[ORF_0622; reading frame open far upstream of start; possible transhilt, linking to previous ORF [Escherichia coli]]	68	55	166
46	9	6866	8015	gi 1155776	[aldehyde dehydrogenase [Vibrio cholerae]]	68	44	1520
48	3	3643	3404	gi 285008	[24k polypeptide [apple stem grooving virus]]	68	47	201
48	4	3536	4132	gi 11045937	[H. genitalium predicted coding region MG246 [Mycoplasma genitalium]]	68	39	597
53	10	11671	10685	gi 1130952	[17qA [Bacillus subtilis]]	68	46	987
70	9	7346	8155	gi 1147198	[lphN protein [Escherichia coli]]	68	40	810
89	4	1899	2566	gi 145173	[35 kDa protein [Escherichia coli]]	68	43	1068
108	1	2187	1150	gi 18722	[precursor (aa -20 to 38) (Acinetobacter calcoaceticus) ir A29277 A29277]	68	57	1038
					aldose-1-epimerase [EC 5.1.3.3] - <i>Acinetobacter lwoffii</i>			
112	5	2666	1622	gi 153724	[MalC [Streptococcus pneumoniae]]	68	55	937
					lipoprotein protein [Bacillus subtilis]			
116	7	7865	8638	gi 143608	[QcrB [Bacillus subtilis]]	68	48	774
118	3	2684	3698	gi 11301805	[QcrB [Bacillus subtilis]]	68	46	1215
120	2	1624	1594	sp P18038 CTSJ_FPI	[SULFITE REDUCTASE (NADPH) FLAVOPROTEIN ALPHA-COMPONENT [EC 1.8.1.2] (SIR-)]	68	45	171
129	1	1	1011	gi 196107	[argininoacidate lyase [Escherichia coli]]	68	50	1011

TABLE 2

5
10
15
20
25
30
35
40
45
50
55

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
132	3	1867	2739	[gi 216267]	[ORF2 [Bacillus megaterium]]	68	48	873
134	2	848	1012	[gi 147545]	[DNA recombinase [Escherichia coli]]	68	50	165
141	2	372	614	[gi 4872116]	[sei (stress inducible protein) [Glycine max]]	68	36	243
149	7	2554	2660	[gi 145774]	[hsp70 protein (dnak gene) [Escherichia coli]]	68	48	195
155	2	1776	1534	[gi 216283]	[ORF1 [Escherichie coli]]	68	36	243
158	3	1826	3249	[gi PA33401 YCH_]	[HYPOTHETICAL 54.3 kD PROTEIN IN ECO-ALKB INTERGENIC REGION.]	68	51	1464
169	6	2749	3118	[gi 1403402]	[Unknown [Mycobacterium tuberculosis]]	68	46	570
175	10	9158	7365	[gi 1072195]	[phaa gene product [Rhizobium meliloti]]	68	51	1794
188	7	4184	5134	[gi 1170843]	[13-ketoacyl-ACP synthase II [Vibrio harveyi]]	68	48	1251
189	1	907	1665	[gi 1467793]	[tRNA binding protein (probable) [Bacillus subtilis]]	68	55	759
206	5	7683	6709	[gi 14250138]	[yebI [Bacillus subtilis]]	68	48	975
206	8	10425	12176	[gi 152887]	[lipoate decarboxylase [Saccharomyces cerevisiae]]	68	48	1752
212	8	3421	3648	[gi 1336941]	[c1 gene product [Bacteriophage B1]]	68	39	226
214	8	5457	6462	[gi 1420467]	[orf Y0194c [Saccharomyces cerevisiae]]	68	45	1026
237	4	2507	3088	[gi 149881]	[HslH [Lactococcus lactis]]	68	46	582
243	5	5540	4542	[gi 223684]	[malonate pyrophosphate decarboxylase [Saccharomyces cerevisiae]]	68	47	999
262	1	3	164	[gi 150374]	[4-oxalocrotonate tautomerase [Pseudomonas putida]]	68	42	162
262	2	1384	1118	[gi 11147244]	[P8R [Enterococcus hirae]]	68	49	867
276	6	3702	3139	[sp PA0501 ABC_E]	[ATP-BINDING PROTEIN ABC (FRAGMENT)]	68	50	564
306	6	6345	5725	[gi 1256117]	[adenine phosphoribosyltransferase [Bacillus subtilis]]	68	53	621
323	1	4599	3850	[gi 467073]	[Unknown [Bacillus subtilis]]	68	45	750
365	6	5017	4838	[gi 1130643]	[T22B3.3 [Caenorhabditis elegans]]	68	45	180
376	2	569	1646	[gi 1277026]	[DAPA aminotransferase [Bacillus subtilis]]	68	51	1094
405	1	1741	972	[gi 1403917]	[ltpB [Bacillus subtilis]]	68	47	910
406	2	853	539	[gi 1511513]	[ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]]	68	44	315
426	6	3558	3391	[gi 622632]	[lgtI [Escherichia coli]]	68	48	168
438	1	108	329	[gi 146923]	[nitrogenase reductase [Escherichia coli]]	68	43	222

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	orf	Start (nt)	Stop (nt)	match addressan	match gene name	% sim	% ident	length (nt)
443	1	476	240	[gi 515810	hippuricase (Campylobacter jejuni)	68	42	237
443	2	518	1015	[gi 1204742	[H_ influenzae predicted coding region H10591 (Haemophilus influenzae)	68	48	498
443	5	4447	3779	[gi 514650	deoxyribose-phosphate aldolase (Bacillus subtilis) pir S49551 S49455	68	55	669
476	2	240	1184	[gi 571345	unknown, similar to E. coli cardiolipin synthase (Bacillus subtilis) np P45860 YNE_BACSU HYPOTHETICAL 38.2 PROTEIN IN NANI-ACDA INTERGENIC REGION.	68	45	945
486	2	1876	1046	[gi 147228	transport protein (Escherichia coli)	68	41	831
517	3	1764	2084	[gi 1523609	[orf2 [Bacteriophage A2]	68	64	321
572	1	2	571	[sp P19227 YOSL_	HYPOTHETICAL 6.8 KD PROTEIN IN MRDC-TK INTERGENIC REGION.	68	47	570
646	1	914	459	[gi 413932	[16S-5S gene product (Bacillus subtilis)	68	52	456
659	3	1668	1391	[gi 1407341	[C3D5_B [Caenorhabditis elegans]	68	36	234
864	1	1510	1716	[gi 145774	[hsp70 protein (dnaK gene) [Escherichia coli]	68	48	207
920	1	860	432	[gi 1510116	[hypothetical protein ISP_P11461 (Methanococcus jannaschii)	68	54	429
952	1	1096	611	[gi 123456	[reductase (Leishmania major)	68	46	486
970	1	91	402	[gi 1354775	[ptosR [Treponema pallidum]	68	46	312
1026	1	1064	534	[gi 1510117	[thiaminopimelate decarboxylase (Bacillus subtilis)]	68	47	531
1029	1	426	216	[gi 1135714	[plasmid parM gene for asparagine-rich antigen (clone 17c1)]	68	31	213
1058	1	692	1348	[gi 1510169	[epic gene product [Staphylococcus epidermidis]	68	46	345
1096	2	665	465	[gi 141434	[rho factor (Bacillus subtilis)]	68	43	201
1308	1	2	694	[gi 16939	[group B oligopeptidase, PBP B [Streptococcus agalactiae]]	68	50	693
1679	1	2	238	[gi 117205	[67 kDa Myo-inositol-crossreactive streptococcal antigen (Streptococcus yoelii)]	68	53	237
2039	1	3	383	[gi 153898	[transport protein (Salmonella typhimurium)]	68	51	381
2077	1	3	126	[pir C3196 C324	[hisC homolog - Bacillus subtilis]	68	47	324
2112	1	613	374	[gi 66884	[lamin L1 (Keropus laevis)]	68	50	240
2273	1	793	398	[gi 581668	[epic gene product [Staphylococcus epidermidis]]	68	45	396
2948	1	2	385	[gi 216869	[branched-chain amino acid transport carrier (Pseudomonas aeruginosa) pir A3853 A48534 branched-chain amino acid transport protein braZ]	68	41	384
					Pseudomonas aeruginosa			

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins5
10
15
20
25
30
35
40
45
50

Contig	ORF	Start	Stop	match accession	match gene name	% sim	% ident	length (nt)
ID	ID	(nt)	(nt)					
2955	1	768	400	gi 204179	[hypothetical protein [Bacillus subtilis]]	68	49	369
2981	1	572	288	gi 508979	[GTP-binding protein [Bacillus subtilis]]	68	48	285
1014	1	584	294	gi 1524394	[ORF-2 upstream of gbsA operon [Bacillus subtilis]]	68	45	291
1082	1	336	169	gi 1204696	[tructose-phosphate IBC component [Haemophilus influenzae]]	68	53	168
3108	1	103	258	gi 212855	[heat-shock protein [Arabidopsis thaliana]]	68	48	156
3639	1	919	461	gi 1520190	[nitrate transport permease protein [Methanococcus jannaschii]]	68	47	459
3657	1	1	330	gi 1553589	[PS enzyme-11 fructose [Xanthomonas campestris]]	68	48	310
3823	1	780	391	gi 603768	[putative protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] [Bacillus subtilis]]	68	54	390
3982	1	2	277	gi 1169435	[putative [Lactococcus lactis]]	68	47	276
4051	1	1	142	gi 450686	[hypothetical gene product [Escherichia coli] pir S3437 S38037] [hsdM protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli [SUB 40-520]]	68	48	342
4089	1	12	209	gi 1153678	[heavy-metal transporting B-type ATPase [Proteus mirabilis]]	68	47	198
4143	1	47	187	gi 603769	[HUU protein, urocanase [Bacillus subtilis]]	68	55	141
4148	1	2	352	gi 450688	[hypothetical gene product [Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli [SUB 40-520]]	68	51	351
4173	1	2	182	gi 1101037	[pyruvate kinase [Bacillus psychrophilus]]	68	48	181
4182	1	1	498	gi 1413968	[lpa-44d gene product [Bacillus subtilis]]	68	50	249
4362	2	148	318	gi 450688	[hypothetical gene product [Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli [SUB 40-520]]	68	44	171
5	111	9493	800	gi 141727	[putative [Bacillus subtilis]]	67	46	1194
31	111	10318	9333	gi 216746	[D-lactate dehydrogenase [Lactobacillus plantarum]]	67	41	486
32	1	1560	1155	gi 1088557	[renal sodium dicarboxylate cotransporter [Homo sapiens]]	67	46	1596
32	5	4945	4145	gi 1510720	[l-phenylalanine dehydrogenase [Methanococcus jannaschii]]	67	51	801
36	5	5350	4268	gi 146216	458 identity with the product of the ORF6 gene from the <i>Erwinia herbicola</i> carotenoid biosynthesis cluster; putative [Bacillus subtilis]	67	58	1083
44	7	4492	5104	gi 11006621	[hypothetical protein [Synechocystis sp. P-1]]	67	43	813

TABLE 2

55

S. aureus - Putative coding regions of novel proteins similar to known proteins

5
10
15
20
25
30
35
40
45
50

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
56	7	3943	8461	[gi 324131	glutamate synthase large subunit precursor (<i>Azospirillum brasiliense</i>)	67	52	4539
				[pir BA6602 BA6602	glutamate synthase (NADPH) (EC 1.4.1.13) alpha chain - <i>Azospirillum brasiliense</i>			
56	12	113923	14678	[gi 1000453	[trrR (<i>Bacillus subtilis</i>)]	67	48	756
62	8	5692	4757	[gi 1123949	[orf3 (<i>Bacillus</i> , C-125, alkali-sensitive mutant 12224, Peptide Nutant, 112 ab)	67	45	336
62	10	17570	6338	[gi 1054655	[Na/H antiporter system (<i>Bacillus alcalophilus</i>)	67	49	1223
99	3	2129	3221	[gi 1204349	[hypothetical protein (GB:Gb D90212.3) (<i>Haemophilus influenzae</i>)]	67	50	1201
102	9	5695	7176	[gi 149432	[putative (<i>Lactococcus lactis</i>)	67	51	1482
103	13	14549	14049	[gi 1408497	[lpdP gene product (<i>Bacillus subtilis</i>)	67	48	501
109	15	114821	113922	[gi 113976	[lpa-scr gene product (<i>Bacillus subtilis</i>)	67	49	840
109	17	114811	15194	[gi 113983	[lpa-s9d gene product (<i>Bacillus subtilis</i>)	67	29	184
121	4	17133	2153	[gi 11262335	[ymA (<i>Bacillus subtilis</i>)	67	54	441
122	3	1	1	[gi 143047	[orbB (<i>Bacillus subtilis</i>)	67	35	119
124	5	4060	3518	[gi 1556885	[unknown (<i>Bacillus subtilis</i>)	67	47	543
121	2	4584	3589	[gi 1046081	[hypothetical protein (GB:D2615_10) (<i>Mycoplasma genitalium</i>)]	67	30	996
140	3	2499	2287	[gi 146549	[kdpc (<i>Escherichia coli</i>)	67	45	603
142	4	5109	4198	[gi 1212775	[GTP cyclohydrolase II (<i>Bacillus mywoliqufaciens</i>)	67	55	1212
147	5	2913	2374	[gi 1303768	[YraJ (<i>Bacillus subtilis</i>)	67	44	540
152	8	6341	6673	[gi 13779841	[unknown (<i>Bacillus subtilis</i>)	67	48	333
161	4	2720	13763	[gi 496319	[spbx (<i>Synechococcus</i> sp.)	67	47	1044
163	6	1989	3428	[gi 595681	[2-oxoglutarate/malate translocator (<i>Spinacia oleracea</i>)	67	47	1440
153	3	1351	1626	[gi 1511101	[shikimate 5-dihydrogenase (<i>methanococcus jannaschii</i>)	67	53	276
200	2	917	2179	[gi 142439	[ATP-dependent nuclelease (<i>Bacillus subtilis</i>)	67	48	1263
206	10	12445	112801	[sp P37347 YEC0_	[HYPOTHETICAL 21.8 kD PROTEIN IN ASPS 5' REGION.	67	47	357
206	11	113047	144432	[gi 1732813	[branched-chain amino acid carrier (<i>Lactobacillus delbrueckii</i>)	67	46	1386
208	2	1321	809	[gi 1033037	[100 kDa heat shock protein (Hsp100) (<i>Leishmania major</i>)	67	36	513
218	3	1019	2052	[gi 109542	[ChrB protein (<i>Erwinia chrysanthemi</i>)	67	43	1014

TABLE 2

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)	
10	246	2	176	gi 215098	excisionase [Bacteriophage 15a]	67	37	192	
1276	2	2260	1412	gi 303560	[ORF27] [Escherichia coli]	67	50	849	
297	6	2223	3056	gi 142784	[C2A protein [Bacillus firmus]	67	46	834	
107	7	5220	4186	gi 1070013	protein-dependent [Bacillus subtilis]	67	43	1035	
316	1	16	1028	gi 1161061	[dioxigenase [Methyllobacterium extorquens]	67	57	991	
324	3	5620	5030	gi 1469784	[putative cell division protein ftsW [Enterococcus hirae]	67	49	621	
336	1	524	264	gi 173122	[urea amidolyase [Saccharomyces cerevisiae]	67	45	261	
310	1	108	1	gi 1070531	[IS1101-TRNA SYNTHETASE [EC 6.1.1.21] [IS1101-TRNA (TIGASI) (HTSRI)]	67	47	1287	
164	3	4890	3592	gi 151259	[NHC-CoA reductase [EC 1.1.1.88] (Pasuomonas mesonilii) pir A4756 A44756	67	46	1299	
165	3	2940	2113	gi 1296223	[hydroxymethylglutaryl-CoA reductase [EC 1.1.1.88] Pseudomonas sp.]	67	47	828	
367	2	325	1	918	gi 1039479	[orf2 gene product [Lactobacillus helveticus]	67	47	594
395	1	666	1221	gi 1204516	[hypothetical protein [GB_00014_4] Haemophilus influenzae]	67	55	606	
415	1	1800	901	gi 822519	[ICG Site No. 29739 [Escherichia coli]	67	46	900	
419	1	1799	903	gi 520752	[putative [Bacillus subtilis]	67	48	897	
474	1	2	796	gi 86906	[argininosuccinate synthetase [Streptomyces clavuligerus] pir S7655 S77659	67	49	795	
485	2	1921	2226	gi 103434	[argininoacetyl synthase [EC 6.3.4.5] - trentoxycytes clavuligerus	67	43	306	
596	1	1728	865	gi 1103053	[Ygf [Bacillus subtilis]	67	47	864	
700	1	433	1216	gi 1204628	[hypothetical protein (SP_P21498) Haemophilus influenzae]	67	47	216	
806	2	249	647	gi 1677947	[AppC [Bacillus subtilis]	67	51	399	
828	2	340	900	gi 177761	[lrrA [Synechococcus sp.]	67	37	561	
833	1	1407	916	gi 142956	[regulatory protein [Bacillus subtilis]	67	41	492	
856	1	1555	779	gi 170224	[25970_2 [Caenorhabditis elegans]	67	38	777	
888	1	1614	850	gi 437315	[m7C start codon [Bacillus licheniformis]	67	40	765	
1034	1	1190	597	gi 1105113	[hypothetical protein (GB_13201_15) Haemophilus influenzae]	67	45	594	
1062	1	616	319	gi 1103050	[Yopc [Bacillus subtilis]	67	41	318	
1067	1	918	460	gi 122950 A329	[probable reductase protein - Leishmania major]	67	54	459	

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	sim	ident	length (nt)	
1358	1	3	293	gi 1601369	[hypothetical protein [Synthecystis sp.]	67	44	291	
2101	1	1	302	gi 1510016	[hypothetical protein [SP:R1466] [Methanococcus jannaschii]	67	48	300	
3000	1	1	507	gi 1517205	[67 kDa Myo-in-crossreactive streptococcal antigen [Streptococcus yoghensis]	67	56	507	
3066	1	464	234	gi 308861	[tRNA start codon [Lactococcus lactis]	67	46	231	
3087	1	454	251	gi 1205166	[oligopeptide transport ATP-binding protein [Haemophilus influenzae]	67	44	204	
3101	1	2	256	gi 1533541	[porphyrinogen III methyltransferase [Sea urchin]	67	55	255	
3598	1	728	393	gi 1512359	[HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas aeruginosa] pJ44756 A44756	67	56	336	
3765	1	2	584	gi 1557469	[hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) [Pseudomonas sp.]	67	45	219	
3788	1	658	398	gi 1552915 S229	[nitrate reductase alpha chain - Bacillus subtilis (fragment)]	67	45	261	
3883	1	2	265	gi 1704397	[cyathionine beta-lyase [Arabidopsis thaliana]	67	46	264	
3926	1	2	340	gi 1461199	[peptide synthetase [Amycolatopsis mediterranei]]	67	44	339	
4417	1	82	396	gi 1205337	[ribonucleotide transport ATP-binding protein [Haemophilus influenzae]]	67	46	315	
1	2	13	3075	gi 155348	[cdv [Bacillus subtilis]]	66	42	915	
15	1	6	2273	gi 12542	[satB [Synthecystis PCC7942]]	66	37	270	
11	1	9	8059	gi 12056	[lacin [Ustilago maydis]]	66	44	234	
31	10	9034	9258	gi 1204345	[mercury scavenger protein [Haemophilus influenzae]]	66	48	229	
32	1	6	6347	gi 15253	[gi 998342	inducible nitric oxide synthase [Gallus gallus]	66	47	1095
44	13	8856	10124	gi 1520751	[molybdenum cofactor biosynthesis mta protein [Methanococcus jannaschii]]	66	46	1269	
48	1	2	1276	gi 12068	[gi 1502099	[orf 1 [Mycoplasma mycolae]]	66	40	1593
58	1	8	7178	gi 1665939	[hypothetical protein [Bacillus subtilis]]	66	47	1251	
62	1	7	5143	gi 1370	[lipA gene product [Rhizobium meliloti]]	66	40	774	
70	14	11683	10938	gi 809660	[deoxyribose-phosphate adolase [Bacillus subtilis] pJ4455 S9455	66	55	696	
76	1	1	1305	gi 142440	[ATP-dependent nuclelease [Bacillus subtilis]]	66	42	1305	
91	1	6	9236	gi 1704397	[cyathionine beta-lyase [Arabidopsis thaliana]]	66	43	1032	
102	1	5	3810	gi 1204323	[hypothetical protein [SP:R1805] [Haemophilus influenzae]]	66	41	546	

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start	Stop	match	match gene name	% sim	% ident	length (nt)
ID	ID	(nt)	(nt)	accession				
103	4	3418	2732	gi 971344	nitrate reductase gamma subunit [Bacillus subtilis] sp P42177 NR1_BACSU	66	48	687
					nitrate reductase gamma chain [EC 1.7.99.4] .gi 0093165			
					nitrate reductase [Bacillus subtilis] (sub -160)			
109	6	4243	4674	gi 170886	glucosamine-6-phosphate deaminase [Candida albicans] pir A6552 A46652	66	45	432
					glucosamine-6-phosphate isomerase			
112	117	117491	117212	gi 1323179	109F YGR11V [Saccharomyces cerevisiae]	66	33	222
116	2	4667	2637	gi 1491813	gamma-glutamyltranspeptidase [Bacillus subtilis]	66	43	2031
150	5	3189	2889	gi 11146224	putative [Bacillus subtilis]	66	30	201
172	5	3864	3662	gi 1755152	highly hydrophobic integral membrane protein [Bacillus subtilis]	66	41	399
					sp P42933 TAGC_BACSU			
					ACID TRANSPORTATION PERMEASE PROTEIN AGC.			
174	5	4592	3723	gi 11146241	panthenate synthetase [Bacillus subtilis]	66	49	870
175	4	3209	2880	gi 1642655	unknown [Rhizobium meliloti]	66	29	330
175	111	8743	7994	gi 1851655	Na/H antiporter system [Bacillus alcelaphilus]	66	43	750
190	5	17079	5727	gi 161072	di-tripeptide transporter [Lactococcus lactis]	66	40	1353
195	115	113919	137713	gi 11322411	unknown [Mycobacterium tuberculosis]	66	42	207
217	3	2822	2595	gi 1114542	alternative stop codon [Rattus norvegicus]	66	36	228
233	9	7113	6135	gi 1458327	F08P3_4 gene product [Cenorhabditis elegans]	66	47	999
248	1	1041	1041	gi 1095941	[C]hA protein [Erwinia chrysanthemi]	66	42	994
241	1	2102	1053	gi 1530667	peptidoglycan hydrolase [Staphylococcus aureus]	66	33	1050
261	1	1178	648	gi 1510859	[H. Jannaschii] predicted coding region H30790 [Methanococcus jannaschii]	66	40	531
1	3731	12973	911205865	[L]actohydrodipicolinate N-succinyltransferase [Haemophilus influenzae]	66	47	759	
272	8	6348	5484	gi 1882101	high affinity nickel transporter [Alcaligenes eutrophus]	66	44	1065
					#gi P23516 IIO20_ALCEU HIGH-AFFINITY NICKEL TRANSPORT PROTEIN.			
276	3	2805	2104	gi 1208965	hypothetical 23.3 kd protein [Escherichia coli]	66	47	702
278	1	2830	1784	gi 1486662	[phosphatase]associated protein [Bacillus subtilis]	66	48	1047
278	3	3830	12932	gi 303560	[ORF271] [Escherichia coli]	66	45	879
279	2	3894	2218	gi 1182789	[2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase [Bacillus subtilis]	66	48	1677
288	4	2535	2275	gi 1256625	putative [Bacillus subtilis]	66	42	261
292	2	1133	942	gi 1511604	[H. Jannaschii] predicted coding region H31651 [Methanococcus jannaschii]	66	30	192

TABLE 2

5
10
15
20
25
30
35
40
45
50

55

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start	Stop	match accession	match gene name	% sim	% ident	length (nt)
1D	1D	(nt)	(nt)					
294	1	1116	559	[gi 216314	[astase [Bacillus stearothermophilus]	66	45	58
297	4	2913	1978	[gi 994794	[cytochrome a assembly factor [Bacillus subtilis] sp P24009 CORK_BACSU	66	45	936
316	4	2051	2662	[gi 1107839	[PROBABLE CYTOCHROME C OXIDASE ASSEMBLY FACTOR.	66	40	630
338	4	2460	2362	[gi 520750	[faginate lyase [Pseudomonas aeruginosa]	66	59	159
339	1	1214	735	[gi 167468	[biotin synthetase [Bacillus sphaericus]	66	52	480
363	1	3	863	[gi 581649	[lipoC gene product [Staphylococcus epidermidis]	66	47	861
366	1	232	483	[gi 110505	[Unknown [Schizosaccharomyces pombe]	66	53	232
367	4	2468	1845	[sp P20692 TMYRA	[PREHENATE DEHYDROGENASE (EC 1.3.1.12) (PDR).	66	50	624
372	3	2150	1599	[gi 467116	[Unknown [Bacillus subtilis]	66	18	532
378	1	212	1009	[gi 147309	[purine nucleotide phosphorylase [Escherichia coli]	66	50	798
401	1	1	462	[gi 388563	[μ -aminobenzoic acid synthase [Streptomyces griseus] pir JN0531 JN0531_P-	66	46	482
404	1	4826	5234	[gi 606744	[α -aminobenzoic acid synthase - Streptomyces faiens	66	51	429
411	2	1728	1103	[gi 1460081	[cytidine deaminase [Bacillus subtilis]	66	44	636
420	1	1	541	[gi 1046024	[Unknown [Mycobacterium tuberculosis]	66	49	540
431	1	1	858	[gi 1500008	[Na ⁺ /ATPase subunit J [Mycoplasma genitalium]	66	10	158
443	7	5679	5299	[gi 852076	[HmgA [Bacillus subtilis]	66	46	381
464	3	3105	2413	[gi 153047	[lysostaphin (LysT) start codon [Staphylococcus simulans] pir J25181 J25181	66	51	933
561	1	956	480	[gi 1204905	[lysostaphin precursor - Staphylococcus simulans sp P10547 LSPSP_STAS1	66	43	714
562	3	1066	1183	[gi 1046082	[lysostaphin PRECURSOR (EC 3.5.1.-1.	66	45	477
576	1	11	724	[gi 1205014	[DNA-3'-methyladenine glycosidase I [Haemophilus influenzae]	66	52	318
577	3	1190	903	[gi 1001153	[H. genitalium predicted coding region HQ372 [Haemophilus genitalium]	66	43	714
584	1	1	2	[gi 21041YBBA	[Hypothetical protein [Synchocystis sp.]	66	52	218
592	1	1410	706	[gi 928339	[HYPOTHETICAL 46.7 KD PROTEIN IN NSBB-RUVB INTERGENIC REGION (ORF1).	66	48	310
601	1	1433	720	[gi 1488695	[ORF266; putative [Lactococcus lactis phage B15-T1]	66	51	705
					[novel antigen; orf-2 [Staphylococcus aureus]	66	55	714

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (Int)	Stop (Int)	match accession	match gene name	% sim	% ident	length (nt)
619	3	468	845	gi 246573	similar to H. musculus transport system membrane protein, Nrpap PIR:AK07391 and S. cerevisiae SNF1 protein (PIR:AK5154) Caenorhabditis elegans	66	45	378
706	2	561	355	gi 604008	[unknown protein (Rattus norvegicus)]	1	66	46
734	2	673	532	gi 1519085	phosphatidylcholine binding immunoglobulin heavy chain IgH variable region	66	60	162
740	1	3	317	gi 120272	[argininosuccinate]lyase [Campylobacter jejuni]	1	66	47
764	1	310	747	gi 435396	[alkaline phosphatase like protein (Lactococcus lactis) pir S59339 S59339]	66	42	418
852	1	338	171	gi 1536955	[alkaline phosphatase-like protein - Lactococcus acacia	66	43	168
886	1	3	158	gi 28972	[tertiochrome-binding protein [Bacillus subtilis]]	66	44	156
889	1	462	232	gi 1033661	[Human cytomegalovirus]	66	66	231
893	1	2	247	gi 149008	[putative (Helicobacter pylori)]	66	45	246
900	1	1425	733	gi 1580442	[F1 [Bacillus subtilis]]	66	51	693
906	2	2300	1473	gi 790345	[acyl-alcohol dehydrogenase [Bacillus subtilis]]	66	53	628
947	1	79	539	gi 410117	[diaminopimelate decarboxylase [Bacillus subtilis]]	66	47	471
950	1	1100	532	gi 48713	[orf145 [Staphylococcus aureus]]	66	35	549
955	1	89	475	gi 11204390	[uridine kinase (uridine monophosphokinase) [Haemophilus influenzae]]	66	50	387
981	2	1108	997	gi 1577446	[ribopyr protein [Plasmodium yoelii]]	66	16	317
986	1	25	315	gi 105002	[orf_F-1556 [Escherichia coli]]	66	31	291
1057	1	3	203	gi 1130853	[YosF [Bacillus subtilis]]	1	66	40
1087	1	1	294	gi 1575313	[Unknown [Saccharomyces cerevisiae]]	66	53	294
1105	1	1	231	gi 1047799	[methylgalactoside permease ATP-binding protein [Mycoplasma genitalium]]	66	46	231
1128	1	2	574	gi 1001493	[hypothetical protein [Synechocystis sp.]]	66	46	573
1150	1	498	250	gi 149034	[H. Jannaschii predicted coding region HJ0355 [Methanococcus jannaschii]]	66	40	249
1180	2	1707	451	gi 1215908	[DNA polymerase (q4) [Bacteriophage T4]]	66	46	255
1208	1	1123	587	gi 1256653	[DNA-binding protein [Bacillus subtilis]]	66	58	537
1342	1	1	402	gi 1120474	[hypothetical protein [Synechocystis sp.]]	66	53	402
1761	2	589	398	gi 1215611	[cell fiber protein [Bacteriophage T5]]	1	66	50

TABLE 2

5
10
15
20
25
30
35
40
45
50
55

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1983	1	499	251	gi 1045935	[DNA helicase II (Helicase II)]	66	40	249
2103	2	176	400	gi 929798	[precursor for the major merozoite surface antigens (Plasmodium vivax)]	66	46	225
2141	1	373	188	gi 1256623	[endoribonuclease (Bacillus subtilis)]	66	38	186
2458	1	135	164	gi 101910	[unknown (Schizosaccharomyces pombe)]	66	47	162
2505	1	466	235	gi 1510394	[putative transcriptional regulator (Methanococcus jannaschii)]	66	39	234
2525	1	558	280	gi 100695	[cytotoxin L (Clostridium sordellii)]	66	44	279
2935	1	3	275	gi 1765033	[antolysin (Staphylococcus aureus)]	66	47	273
3005	1	114	305	gi 1205784	[heterocyst maturation protein (Hemophillus influenzae)]	66	46	192
3048	1	80	277	gi 1309813	[Yew (Bacillus subtilis)]	66	42	198
3071	1	1	189	gi 1070014	[protein-dependent (Bacillus subtilis)]	66	41	189
3081	1	404	225	gi 984212	[unknown (Schizosaccharomyces pombe)]	66	44	180
3090	2	580	386	gi 1204987	[DNA polymerase III, alpha chain (Haemophilus influenzae)]	66	48	195
3118	1	1	387	gi 100366	[respiratory nitrate reductase (Bacillus subtilis)]	66	49	387
3739	1	798	400	gi 110684	[ProV (Bacillus subtilis)]	66	47	399
3796	1	402	202	gi 853760	[acyl-CoA dehydrogenase (Bacillus subtilis)]	66	60	201
1924	1	397	1397	gi 36352	[galactonate permease (Bacillus licheniformis)]	66	46	249
4240	1	3	350	gi 15159	[lnc-CoA reductase (EC 1.1.1.86) Pseudomonas mevalonii] pir AA4756 AA4756	66	31	349
					[hydroxymethylglutaryl-CoA reductase (EC 1.1.1.86) Pseudomonas sp.]			
4604	1	7	234	lpri A2713 BHC	[hemocyanin subunit II - Atlantic horseshoe crab]	66	46	228
4	9	8845	9750	gi 145616	[cymR (Escherichia coli)]	65	35	906
6	5	2708	3565	gi 15159	[lnc-CoA reductase (EC 1.1.1.86) Pseudomonas mevalonii] pir AA4756 AA4756	65	47	658
13	1	1993	998	gi 143102	[recombination protein (ttg start codon) Bacillus subtilis] gi 103922 RecN	65	44	996
15	17	2493	3524	gi 1403126	[cycD gene product (Alcaligenes eutrophus)]	65	38	1032
18	3	1908	1372	gi 349187	[acyltransferase (Saccharomyces cerevisiae)]	65	50	537
21	3	1467	2492	gi 149518	[phosphorylboyl antranilate transferase (Lactococcus lactis)]	65	52	1026
					[pir S15126 S35126 antranilate phosphorylboyltransferase (EC A.2.1.16) - Lactococcus lactis subsp. lactis]			
25	4	3374	4312	gi 150420	[fatty-acyl carrier protein transacylase (Bacillus subtilis)]	65	44	939

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

5
10
15
20
25
30
35
40
45
50

GeneID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt.)
27	2	390	626	[gi 1212129]	[ygbj] [Bacillus subtilis]	65	45	237
31	12	11040	110387	[gi 509215]	[D-hydroxyisocaproate dehydrogenase [Lactobacillus delbrueckii]	65	41	654
38	124	19172	19588	[gi 5477]	[H-protein [Pleurotus croculellus]]	65	41	357
44	2	790	1746	[gi 405882]	[yalk] [Escherichia coli]	65	46	957
44	12	9156	8822	[gi 120505]	[molybdenum cofactor biosynthesis protein [Haemophilus influenzae]]	65	50	525
45	6	6635	7588	[gi 493014]	[Aba protein [Salmonella typhimurium]]	65	46	954
51	2	580	1503	[gi 560897]	[OppB gene product [Bacillus subtilis]]	65	45	924
52	1	225	953	[gi 1205318]	[NAD(P)H-flavin oxidoreductase [Haemophilus influenzae]]	65	45	729
55	4	1139	1058	[pir A4459_A444]	[trponin T beta TnT-5 - rabbit]	65	41	282
67	9	7421	8272	[gi 141607]	[sporulation protein [Bacillus subtilis]]	65	42	852
73	5	4416	5375	[gi 1204896]	[lysophospholipase II [Haemophilus influenzae]]	65	37	930
74	1	954	478	[gi 1204844]	[H. influenzae predicted coding region H10594 (Haemophilus influenzae)]	65	50	477
77	1	2	757	[gi 1046082]	[M. genitalium predicted coding region M0372 (Mycoplasma genitalium)]	65	46	756
77	2	795	1433	[gi 1322116]	[paramase [Haemophilus influenzae]]	65	37	639
81	3	4728	3454	[gi 1001708]	[hypothetical protein [Synechocystis sp.]]	65	49	1275
91	7	6548	6357	[gi 1399263]	[cystathione beta-lyase [Enterococcus nivalans]]	65	40	192
98	1	1608	1988	[gi 467423]	[Unknown [Bacillus subtilis]]	65	38	381
98	1	2250	2987	[gi 467424]	[Unknown [Bacillus subtilis]]	65	45	738
102	3	2596	2119	[gi 1515332]	[N-terminal acetyltransferase complex, subunit AND1 [Inchanococcus jannaschii]]	65	39	480
102	4	13647	2862	[gi 1204637]	[H. influenzae predicted coding region H10588 (Haemophilus influenzae)]	65	32	786
103	9	10831	9841	[gi 142695]	[S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase Bacillus megaterium]	65	47	1011
103	10	10139	10119	[gi 710021]	[Intrite reductase (nicD) [Bacillus subtilis]]	65	51	321
106	2	262	1140	[gi 39881]	[ORF 311 (AA 1-311) [Bacillus subtilis]]	65	44	879
109	5	3909	4268	[gi 1204399]	[glucosamine-6-phosphate deaminase protein [Haemophilus influenzae]]	65	44	360
109	10	7165	8595	[gi 5166955]	[ICG Site No. 311 [Escherichia coli]]	65	41	1431

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	OKF	Start	Stop	match	match gene name	t sim	t ident	length (nt)
ID	ID	Int	Int	accession				
110	4	1688	3915	gi 407861	Stress response-like protein [Streptococcus emm151iis] pir S3975 S3975 stringent response like protein - Streptococcus quinismilis	65	45	228
110	5	3882	4295	gi 4078680	[ORF1] Streptococcus equisimilis	65	50	414
110	6	4231	4380	gi 11139574	[orf2] Streptomyces griseus	65	56	150
112	110	9216	8640	gi 1204571	[h] influenzae predicted coding region H0118 [Haemophilus influenzae]	65	52	579
112	112	112049	112288	gi 1710496	transcriptional activator protein [Bacillus brevis]	65	32	762
125	1	2	202	gi 115158	repeat organellar protein [Plasmodium chabaudi]	65	39	201
126	1	3	422	gi 37589	precursor [Homo sapiens]	65	46	420
127	111	10733	12658	gi 1064809	homologous to sp NTR4_ECOLI [Bacillus subtilis]	65	41	1926
143	8	7543	7004	gi 216513	mutator mutC (nt-GC transversion) [Escherichia coli]	65	56	540
145	5	3587	3818	gi 1209768	[D02_0r1569] [Mycoplasma pneumoniae]	65	27	252
150	4	3482	2841	gi 1116225	[putative] [Bacillus subtilis]	65	37	642
166	1	3858	1948	gi 148304	beta-1,4-N-acetylmuramoylhydrolase [Enterococcus hirae] pir A42296 A42296	65	50	1911
188	6	3195	4178	gi 151943	[lysozyme 2 (EC 3.2.1.-) precursor - Enterococcus hirae (ATCC 9750)]	65	46	984
189	9	4982	4185	gi 154812	[ORF1] putative [Rhodobacter capsulatus]	65	40	198
195	6	7900	5272	gi 1145220	[alanyl-tRNA synthetase [Escherichia coli]]	65	49	2637
195	7	10599	8104	gi 1682711	[exonuclease V alpha-subunit [Escherichia coli]]	65	38	2496
206	16	16896	1891	gi 1408115	[ornithine acetyltransferase [Bacillus subtilis]]	65	53	1296
217	4	3844	3215	gi 1209974	[5'-guanylate kinase [Haemophilus influenzae]]	65	41	630
220	4	5265	3751	gi 580920	[rod (graft) polypeptide (AA 1-673) [Bacillus subtilis] pir S0604 S0604	65	40	1515
					probable rotavirus - bacillus subtilis sp P13484 P13484 PROBABLE POLY(GLYCEROL-PHOSPHATE) LIPID-GLUCOSYLTRANSFERASE (EC 2.4.1.52) [TECHIOLIC ACID BIOSYNTHESIS, ROTAVIRUS E1]			
236	5	2327	3709	gi 1146200	[DNA or RNA helicase, DNA-dependent ATPase [Bacillus subtilis]]	65	46	1341
237	3	1902	2513	gi 149379	[Hs8d] [Lactococcus lactis]	65	46	612
241	4	4968	4195	gi 111205108	[ribonuclease H11 (EC 31264) [Haemophilus influenzae]]	65	50	774
252	1	1278	940	gi 1120989	[hypothetical protein (GB_000022_91) [Haemophilus influenzae]]	65	40	339
261	5	4780	3794	gi 145927	[fcd] [Escherichia coli]	65	43	987

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start	Stop	match	match gene name	% sim	% ident	length
ID	ID	(nt)	(nt)	accession				(nt)
274	1	3	218	[gi 149558]	[Bacillus subtilis] [Bacillus subtilis]	65	42	276
301	2	982	815	[gi 167418]	[Bacillus subtilis] [Bacillus subtilis]	65	45	168
307	4	3586	2664	[gi 1070014]	[protein-dependent (Bacillus subtilis)]	65	40	723
335	2	2286	1399	[gi 146913]	[N-acetylglucosamine transport protein (Escherichia coli B28895) (ECN)] [Escherichia coli sp B09221 PTAA_ECOLI PTS SYSTEM, N-acetylglucosamine-specific IIAcB component YELA]	65	50	888
338	5	4120	3170	[gi 1277029]	[biotin synthase (Bacillus subtilis)]	65	49	951
343	1	1490	2600	[gi 143264]	[membrane-associated protein (Bacillus subtilis)]	65	48	1311
344	4	2761	2531	[gi 11050540]	[tRNA-glycine synthetase (Lupinus luteus)]	65	34	231
358	1	3421	3621	[gi 116220]	[NAD ⁺ -dependent glycerol-3-phosphate dehydrogenase (Bacillus subtilis)]	65	47	201
364	1	238	699	[gi 1130128]	[ORF1 (Staphylococcus aureus)]	65	51	462
379	1	1	576	[gi 1143331]	[alkaline phosphatase regulatory protein (Bacillus subtilis)] [par A27650 A27650 regulatory protein phoA - Bacillus subtilis] [sp P23545 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN NOR (EC 2.7.3.-)]	65	40	576
379	1	1666	1346	[gi 1142268]	[dihydrodolopamide transsuccinylase (cdhB; EC 2.3.1.61) (Bacillus subtilis)]	65	50	681
428	1	187	483	[gi 1420465]	[ORF YOR195W (Saccharomyces cerevisiae)]	65	45	297
438	1	272	138	[gi 14498]	[dnS protein (Bacillus subtilis)]	65	1W	567
444	11	9280	10215	[gi 1204736]	[ribokinase (Influenza)]	65	47	936
449	1	1241	1533	[gi 159848]	[Na/H antiporter homolog (Lactococcus lactis)]	65	41	291
478	1	1452	865	[gi 1105942]	[glycyl-tRNA synthetase (YycD) (central)]	65	39	548
479	1	1032	517	[gi 148192]	[putative (Pseudomonas aeruginosa)]	65	40	516
480	6	4312	5637	[gi 1415662]	[UDP-N-acetylglucosamine 1-carboxyvinyl transferase (Acinetobacter alcoaceticus)]	65	46	1316
484	1	2	430	[gi 146551]	[transmembrane protein (YdpD) (Escherichia coli)]	65	44	429
499	1	54	932	[gi 160456]	[reductase (Lachnmania major)]	65	53	1079
505	1	914	659	[gi 1518853]	[lata (Salmonella typhimurium)]	65	39	456
571	2	1509	683	[gi 149399]	[open reading frame upstream gne (Escherichia coli) ir S37754 S37754] [hypothetical protein XE (Gene 5' region) - Escherichia coli]	65	44	627
611	2	506	270	[gi 10961]	[Rap-2 (Plasmidium falcatum)]	65	40	237

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
705	1	564	283	[gi 710020	[nitrite reductase [nirB] [Bacillus subtilis]	65	52	282
712	1	1	177	[gi 1289272	[ferrichrome-binding protein [Bacillus subtilis]	65	37	177
743	1	2	631	[gi 310631	[ATP-binding protein [Streptococcus gordonii]	65	37	159
749	2	393	779	[gi 1467274	[single strand DNA-binding protein [Bacillus subtilis]	65	45	610
762	1	1688	850	[gi 1160399	[multidrug resistance protein [Plasmodium falciparum]	65	29	387
788	1	85	315	[gi 11225096	[unknown protein [Bacillus sp.]	65	35	221
850	1	1	408	[gi 1006604	[hypothetical protein [Synecystis sp.]	65	37	408
908	1	1	434	[gi 1199346	[2362 [Saccharomyces cerevisiae]	65	46	444
925	1	1	174	[gi 1256553	[DNA-binding protein [Bacillus subtilis]	65	54	174
1031	1	26	232	[gi 2288657	[ADP-Cytochrome d oxidase, subunit 1 homolog [Escherichia coli, K12, epit. 514 aa]	65	47	207
1037	1	414	262	[gi 1491813	[gamma-glutamyltranspeptidase [Bacillus subtilis]	65	46	153
1053	1	348	175	[gi 642655	[unknown [Rhizobium meliloti]	65	34	174
1149	1	1399	752	[gi 1165980	[ribulose-5-phosphate 3-epimerase [Spinacia olaracea]	65	46	648
1214	1	881	495	[gi 1205959	[lactam utilization protein [Haemophilus influenzae]	65	45	387
1276	1	476	276	[pir S15493 S154	[site-specific DNA-methyltransferase S151 (EC 2.1.1.1) - <i>Staphylococcus</i> sanguis	65	35	201
1276	2	900	577	[gi 433794	[ORF [Escherichia coli]	65	34	324
2057	1	272	138	[gi 633699	[Trh3 [Verinia entercoagolica]	65	21	135
2521	1	316	169	[gi 1045789	[hypothetical protein [GB.U4001_76] [Mycoplasma genitalium]	65	41	168
2974	1	590	297	[gi 152052	[enolomerase-selective amidase [Rhodococcus sp.]	65	45	294
3031	1	106	154	[pir JQ1024 JQ10	[hypothetical 30k protein (JW1410 5' region) - fruit fly (Drosophila melanogaster)]	65	45	153
3069	1	3	278	[gi 144906	[product homologous to E. coli thioredoxin reductase; J. Biol. Chem. 1988; 263:9015-9019, and to F52a protein of a 5k hydroperoxide eductase from <i>S. bryophilum</i> ; J. Biol. Chem. (1990) 265:10535-10540; pan reading frame A]	65	46	276
3146	1	282	142	[gi 19315	[ORF gene product [Bacillus subtilis]	65	47	141
3170	1	679	341	[gi 150771	[Indolepyruvate decarboxylase (Erwinia herbicola)]	65	44	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

5
10
15
20
25
30
35
40
45
50
55

Contig	ORF ID	Start (int)	Stop (int)	match accession	match gene name	% sim	% ident	length (nt)
3546	1	1	103	gi 450686	hsdh gene of Ecopri gene product [Escherichia coli] pir S38437 S38447 hsdH	65	42	303
					protein - Escherichia coli pir S09629 S09629 hypothetical protein A -			
3782	1	2	128	gi 166612	[NADH-glutamate synthase [Meditrago sativa]]	65	42	327
3990	1	374	189	gi 1009366	Respiratory nitrate reductase [Bacillus subtilis]	65	53	186
4012	1	613	308	gi 1323127	[ORF YGR087c [Saccharomyces cerevisiae]]	65	50	306
4278	2	726	264	gi 11197667	[vitellogenin [Anolis pulchellus]]	65	42	363
19	4	4259	5518	gi 145727	[dead [Escherichia coli]]	64	45	1260
19	6	7639	6926	gi 11016222	[ycg27 gene product [Cyanophora paradoxa]]	64	36	714
20	8	7053	6454	gi 1765073	[autolyin [Staphylococcus aureus]]	64	47	600
31	13	12706	11537	gi 141009	[ipa-83d gene product [Bacillus subtilis]]	64	45	1170
33	4	2388	4364	gi 1204636	[fructose permease IIBC component [Haemophilus influenzae]]	64	47	1977
36	3	1871	3013	gi 1290503	[glutamate permease [Escherichia coli]]	64	40	1143
37	6	4065	4409	gi 139815	[orf 2 gene product [Bacillus subtilis]]	64	46	345
45	9	7852	8760	gi 1220565	[nucleotide sugar epimerase [Vibrio cholerae O139]]	64	53	909
53	3	1540	1899	gi 11303961	[rqq1 [Bacillus subtilis]]	64	50	360
56	6	4793	3855	gi 457514	[91CC [Bacillus subtilis]]	64	45	919
56	124	30002	30247	gi 170331	[similar to zinc fingers [Catenorhabditis elegans]]	64	42	246
62	4	2759	12421	gi 1642655	[unknown [Rhizobium meliloti]]	64	28	339
85	6	7178	6027	gi 1457702	[5-aminoimidazole ribonucleotide-carboxylase [Pichia methanotolica] pir S3912 S3912 phosphoribosylimidazole carboxylyase [EC 1.1.1.21] - yeast [Pichia methanotolica]]	64	46	1152
96	9	9251	10030	gi 1511513	[ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]]	64	42	780
100	1	1	600	gi 1765073	[autolysin [Staphylococcus aureus]]	64	44	600
106	5	3868	4854	gi 1466778	[lysine specific permease [Escherichia coli]]	64	46	987
123	2	838	554	gi 467484	[unknown [Bacillus subtilis]]	64	47	285
127	8	7514	7810	gi 210061	[serotype-specific antigen [African horse sickness virus] pir S27691 S27891 capsid protein VP2 - African horse sickness virus]	64	26	297
131	7	7134	6721	gi 1511160	[H. jannaschii Predicted coding region K11163 [Methanococcus jannaschii]]	64	46	414

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
142	5	5455	4817	gi 1173517	lipoflavin synthase alpha subunit [Actinobacillus pleuropneumoniae]	64	44	619
143	1	709	356	gi 39501429	[probable reductase protein - Leishmania major	64	52	354
149	10	3555	3295	gi 1398151	[major surface antigen Hs62 (Pneumocystis carinii)]	64	44	261
154	4	3134	2307	gi 9846587	[D1NP [Escherichia coli]]	64	50	828
161	5	3855	480	gi 1903104	[orf72 [Bacillus subtilis]]	64	37	1026
165	1	33	791	gi 467483	[Unknown [Bacillus subtilis]]	64	38	759
175	6	6355	474	gi 107398	[phd gene product [Rhizobium meliloti]]	64	42	1512
188	3	2042	2500	gi 1001961	[HNC class II analog [Staphylococcus aureus]]	64	45	459
195	14	13667	13446	gi 386890	[no definition line found [Escherichia coli]]	64	47	222
206	15	16429	16938	gi 1304134	[argC [Bacillus stearothermophilus]]	64	49	510
215	1	560	282	gi 142359	[orf 6 [Azotobacter vinelandii]]	64	39	279
243	7	7818	6938	gi 414014	[lps-9d gene product [Bacillus subtilis]]	64	49	891
258	2	1130	845	gi 664754	[P17 [Listeria monocytogenes]]	64	38	486
259	1	462	232	gi 149663	[M. jannaschii predicted coding region MJ0837 [Methanococcus jannaschii]]	64	52	231
263	6	6565	5567	gi 142028	[aspartate semialdehyde dehydrogenase [Bacillus subtilis]]	64	48	999
				sp Q06797 M1AS_BACSU	[ASPARTATE-SEMIALDEHYDE DEHYDROGENASE (EC 2.1.1.11)]			
					(ASA DEHYDROGENASE),			
271	1	3	1163	gi 1467991	[hflM; B2235_C2_202 [Mycobacterium leprae]]	64	44	1161
280	1	173	1450	gi 1303839	[yqrH [Bacillus subtilis]]	64	43	1278
293	1	2532	1267	gi 1147245	[prokaryomal protein n' [Escherichia coli]]	64	45	1266
295	2	742	1488	gi 459266	[potential membrane spanning protein (Staphylococcus hominis)]	64	39	747
				pir S12932 S42932	[potential membrane spanning protein - taphylococcus hominis]			
301	5	1625	1446	gi 580035	[lysine decarboxylase [Bacillus subtilis]]	64	35	180
315	4	5064	3949	gi 143396	[quinol oxidase [Bacillus subtilis]]	64	45	1116
321	1	1264	635	gi 71096	[transcriptional activator protein [Bacillus brevis]]	64	41	630
333	5	4520	4339	gi 1311295	[orf2; putative 19 kDa protein [Listeria monocytogenes]]	64	43	282
342	1	1	159	gi 142340	[ftsA [Bacillus subtilis]]	64	38	549
353	3	2878	2124	gi 517049	[ORF_o470 [Escherichia coli]]	64	44	555

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (Incl)	Stop (Incl)	match accession	match gene name	t sim	t ident	length (nt)
379	2	827	1658	Ipir S25295 A328	lipoxygenase (lipoamida) (EC 1.2.4.2) - <i>Bacillus subtilis</i>	64	47	2432
404	6	4429	4819	Ipir A36933 A359	diacylglycerol kinase homolog - <i>Streptococcus mutans</i>	64	35	411
407	1	2020	1133	I91 96926	[orfX] (<i>Bacillus subtilis</i>)	64	41	688
425	1	1109	591	I91 1146177	[phosphotransferase system glucose-specific enzyme II (<i>Bacillus subtilis</i>)]	64	44	519
443	6	4082	4798	I91 147309	purine nucleoside phosphorylase [<i>Escherichia coli</i>]	64	51	717
450	2	1035	1604	I91 04376	[orfT_0162] (<i>Escherichia coli</i>)	64	38	570
470	5	1660	6107	I91 116948	host interacting protein (<i>Bacteriophage B1</i>)	64	45	4428
486	4	1911	1471	I91 205582	[spermidine/putrescine transport system permease protein [<i>Haemophilus influenzae</i>]]	64	35	441
497	1	2217	1159	I91 P6929 PNU_E	[fhu protein]	64	38	1059
501	1	3	410	I91 142450	[ahc protein (<i>Bacillus subtilis</i>)]	64	38	408
514	1	3	290	I91 120496	[H. influenzae predicted coding region H1023B (<i>Haemophilus influenzae</i>)]	64	34	288
551	4	3162	3223	I91 1204511	bacterioferritin conjugatory protein [<i>Haemophilus influenzae</i>]	64	41	162
603	4	759	956	I91 755823	[FADH dehydrogenase F (<i>Streptogyrus asparticanus</i>)]	64	35	198
653	2	940	746	I91 2213234	[dicarboxylic amino acids DdSpP permease (<i>Saccharomyces cerevisiae</i>)]	64	41	195
660	1	3001	2257	I91 P66133 YIAH	[hypothetical protein in OCT 5 region (fragment)]	64	39	1345
695	1	11	502	I91 1001383	[hypothetical protein (<i>Synechocystis sp.</i>)]	64	41	692
702	1	3	752	I91 14285	[DNA primase (<i>Bacillus subtilis</i>)]	64	46	750
826	1	1	339	I91 9271336	[arginyl tRNA synthetase (<i>Bacillus subtilis</i>)]	64	50	339
838	1	1831	917	I91 1354775	[pirS8A (<i>Treponema pallidum</i>)]	64	41	915
864	3	675	944	I91 39833	cyclomaltodextrin glucanotransferase (<i>Bacillus stearothermophilus</i>) 1 39835	64	47	270
887	1	3	677	I91 153002	enterotoxin type E precursor (Staphylococcus aureus) pir A2A179 A28179	64	46	675
928	2	1172	963	I91 311976	enterotoxin type E precursor - Staphylococcus aureus sp P1299 ETTE_STAU	64	42	195
1049	2	800	606	I91 1049115	fibrinogen-binding protein (Staphylococcus aureus) pir S34270 S34270	64	41	210
1067	2	999	748	I91 1151072	fibrinogen-binding protein - Staphylococcus aureus pir P660 Bacillus subtilis	64	50	252

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)	
1120	1	50	202	gi 142439	[ATP-dependent nuclease (Bacillus subtilis)]	64	30	153	
1125	1	751	377	gi 581648	[leuB gene product (Staphylococcus epidermidis)]	64	44	375	
1686	1	402	214	gi 10116517WNS	[transforming protein K-ras - mouse]	64	47	189	
2472	1	2	358	gi 487282	[Hsp - ATPase subunit 3 (Enterococcus hirae)]	64	36	357	
2989	1	520	356	gi 304134	[argC (Bacillus stearothermophilus)]	64	50	165	
3013	1	610	152	gi 1551699	[cytochrome oxidase subunit 1 (Bacillus firmus)]	64	51	279	
3034	1	546	274	gi 11201449	[hypothetical protein (GB:090212.3) (Haemophilus influenzae)]	64	50	273	
3197	1	613	308	gi 1093166	[respiratory nitrate reductase (Bacillus subtilis)]	64	46	306	
3303	1	90	362	gi 1107839	[alginic lyase (Pseudomonas aeruginosa)]	64	43	273	
3882	1	82	288	gi 1215746	[D-lactate dehydrogenase (Lactobacillus plantarum)]	64	42	207	
3888	1	1	312	gi 1149435	[putative (Lactococcus lactis)]	64	48	312	
3918	1	660	331	gi 15532	[acyl-CoA acyltransferase (Yarrowia lipolytica)]	64	46	330	
4000	1	112	378	gi 194688	[Unknown (Saccharomyces cerevisiae)]	64	44	267	
4009	1	81	368	gi 139372	[grB gene product (Bacillus brevis)]	64	41	288	
4166	1	2	149	gi 1149435	[putative (Lactococcus lactis)]	64	46	248	
4366	1	2	307	gi 1216267	[ORF2 (Bacillus megaterium)]	64	43	306	
4457	1	2	400	gi 11197657	[vitellogenin (Anolis pulchellus)]	64	43	399	
11	3	1539	2438	gi 1430228	[orf C (Staphylococcus aureus)]	63	32	900	
24	7	5611	5423	gi 1369943	[al gene product (Bacteriophage 81)]	63	34	189	
29	1	1	390	gi 464441	expressed at the end of exponential growth under conditions in which he enzymes of the TCA cycle are repressed (Bacillus subtilis) ORF (46740) expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed BacC1				
31	6	6329	5712	gi 436943	[ORF (Saccharomyces cerevisiae)]	63	43	390	
44	123	14659	115019	gi 10444610QC	[hypothetical protein F-92 - Escherichia coli]	63	36	351	
48	6	4403	6250	gi 43498	[pyruvate synthase (Halobacterium halobium)]	63	42	1848	
50	5	3869	4738	gi 413967	[lpaA-1d gene product (Bacillus subtilis)]	63	43	870	
53	6	6764	5742	gi 1474176	[regulator protein (Staphylococcus xylosus)]	63	49	1023	

TABLE 2

5
10
15
20
25
30
35
40
45
50

3. *luteus* - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
56	14	15880	112607	[gi 46409	[DNA polymerase III subunit [Bacillus subtilis]	63	44	1728
57	11	7945	7376	[gi 17036	[ORF 0158 [Escherichia coli]	63	39	570
62	3	2419	2114	[gi 142656	[Unknown [Rhizobium meliloti]	63	42	166
70	8	6562	7353	[gi 1399821	[PHOC [Rhizobium meliloti]	63	46	792
75	2	223	927	[gi 169376	[HsG [Lactococcus lactis]	63	45	705
78	5	4912	4401	[gi 413930	[Itp26d gene product [Bacillus subtilis]	63	42	510
91	5	9076	7220	[gi 466997	[IacH2; B2126-C1_157 [Mycobacterium leprae]	63	41	1657
91	8	10566	9448	[gi 1206344	[Cystathione gamma-synthase [Haemophilus influenzae]	63	45	1119
120	1	21	1508	[gi 802657	[Sulfite reductase [NAOPH] flavoprotein beta subunit [Escherichia coli]	63	46	1488
120	4	2722	4125	[gi 665994	[Hypothetical protein [Bacillus subtilis]	63	34	1404
127	7	6064	7566	[gi 40162	[invE gene product [Bacillus subtilis]	63	44	1503
149	6	2321	2106	[gi 1148503	[dnak [Erysipelothrix rhusiopathiae]	63	40	216
149	26	10445	10170	[gi 4870	[ORF 2, has similarity to DNA polymerase [Saccharomyces kluyveri] plasmid pSKL	63	42	276
164	2	507	1228	[gi 145476	[ICP-diglyceride synthetase [Escherichia coli]	63	44	792
166	6	1909	1644	[gi 151912	[Fructose-1,6-bisphosphate phosphotransferase [Rhodobacter capsulatus]	63	41	1746
169	4	1704	1886	[gi 152886	[Elongation factor Ts (tsf) [Spiroplasma citril]	63	48	183
188	5	3145	2951	[gi 1114547	[CIV COI 114 grp IB protein [Podospora anserina]	63	42	195
195	113	11767	112804	[gi 606100	[ORF_035 [Escherichia coli]	63	40	1038
201	2	607	2283	[gi 433534	[Arginyl-tRNA synthetase [Corynebacterium glutamicum p[499]6 [A499]6	63	46	1677
206	14	15893	164689	[gi 560828	- ormybacterium glutamicum			
220	5	7769	5766	[gi 216334	[AsCA protein [Bacillus subtilis]	63	49	597
221	1	74	907	[gi 677945	[AppA [Bacillus subtilis]	63	42	2004
227	3	944	1768	[gi 11510558	[Cobyrin acid synthase [Methanococcus jannaschii]	63	42	834
261	2	804	1070	[gi 486511	[ORF YKR054c [Saccharomyces cerevisiae]	63	46	765
269	2	3606	1960	[gi 48221	[DNA-dependent ATPase, DNA helicase [Escherichia coli] p[501]7 [IVECRO recQ protein - Escherichia coli	63	42	1647

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	sim	ident	length (nt)
278	8	7417	6176	[gi 699273]	cytathionine gamma-synthase [Mycobacterium leprae] sp P46807 NETB_MYCLE CYSATHIONINE GAMMA-SYTHASE [EC 4.2.99.9] O-SUCCINYLHOMOCYSTEINE (THIOL)-LYASE	63	41	1242
287	1	2	738	[gi 1405133]	putative [Bacillus subtilis]	63	38	996
295	1	2	748	[gi 1239983]	hypothetical protein [Bacillus subtilis]	63	41	747
326	3	2148	3134	[gi 45302]	carrier protein (AA 1 - 437) [Pseudomonas aeruginosa] sp S11497 S11497 branched-chain amino acid L-tyrosine [Bacillus subtilis]	63	36	947
362	1	2	1216	[sp P25136 SERP_1]	D-3-PHOSPHOGLYCERATE DEHYDROGENASE [EC 1.1.1.95] (PCD1)	63	38	411
404	1	326	1051	[gi 1303816]	[YqoZ] [Bacillus subtilis]	63	35	726
405	1	3	2101	[gi 1215]	[Yqhv] [Bacillus subtilis]	63	42	387
406	1	451	227	[gi 112152]	sulfate permease (919 start codon) [Synechococcus PCC6310] sp A30101 GYC57 sulfate transport protein - Synechococcus sp. PCC 7942	63	43	235
415	1	2	1048	[gi 1205402]	[Transport ATP-binding protein [Haemophilus influenzae]]	63	41	1671
426	4	3575	2679	[gi 393268]	[29-kilodalton protein [Streptococcus pneumoniae]] sp P42362 P29K_STRNP 29 KD	63	39	897
505	1	3	1347	[gi 1418999]	[membrane protein in PSAA 5 REGION ORF1]	63	40	849
507	1	2	574	[gi 546917]	[comK] [Bacillus subtilis, E26, Peptide, 192 aa]	63	35	573
562	1	2	146	[gi 1084]	[lfts-like gene [lactobacillus delbrueckii]	63	45	939
675	1	1	427	[gi 1510994]	[serine aminotransferase [Methanococcus jannaschii]]	63	29	213
686	1	1	3	[gi 517356]	[nitrate reductase [NADH] [Lotus japonicus]]	63	52	228
701	1	1	3	[gi 801940]	[NQRQ protein [Paracoccus denitrificans]]	63	41	390
720	1	1	2	[gi 47166]	[open reading frame [streptomyces lividans]	63	35	399
779	1	1	571	[gi 1263932]	[unknown [Mycobacterium tuberculosis]]	63	41	285
907	1	1	22	[gi 149445]	[ORF1] [Lactococcus lactis]	63	27	300
972	1	794	199	[gi 1511235]	[H. Jannaschii predicted coding region AJ1232 [Methanococcus jannaschii]]	63	27	396
1085	1	1	1154	[gi 1204277]	[hypothetical protein [GBU00019_14] [Haemophilus influenzae]]	63	38	537
1094	1	1	3	[gi 790943]	[urea amidolyase [Bacillus subtilis]]	63	39	540
1108	1	1	482	[sp S49692 S498]	[regulation protein - Bacillus subtilis]	63	44	480
1113	1	1	1221	[gi 1493017]	[endocarditis specific antigen (Enterococcus faecalis)]	63	45	615

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (Int)	Stop (Int)	match accession	match gene name	% sim	% ident	length (nt)
<i>S. aureus</i> - Putative coding regions of novel proteins similar to known proteins								
1300	1	3	695	lsp P33940 YQ0H_	[HYPOTHETICAL 54.3 KD PROTEIN IN ECO-ALKB INTERGENIC REGION]	63	46	693
1325	1	1	204	lgi 924889	[alp0 protein [Borrelia burgdorferi]]	63	30	204
1814	1	3	245	lgi 1305914	[lghy (Bacillus subtilis)]	63	34	243
2021	1	498	250	lpir C33496 C334	[hsc homolog - Bacillus subtilis]	63	46	249
2325	1	2	193	lgi 436132	[product is similar to TopA of transposon Tn554 from <i>Staphylococcus aureus</i>]	63	40	192
2335	1	1	195	lgi 1184298	[flagellar Hs-ring protein [Borrelia burgdorferi]]	63	47	195
2406	1	451	237	lgi 1041785	[lthopty protein (Plasmodium yoelii)]	63	33	225
2961	2	136	360	lgi 1312443	[carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldoliticus]]	63	52	225
2965	1	1	402	lgi 1407784	[orf-1; novel antigen [Staphylococcus aureus]]	63	50	402
2987	1	501	293	lgi 11224069	[lmidase [Mycobacteria cathephalisa]]	63	35	291
2994	1	266	135	lgi 836846	[lphosphoribosylformamino-peptidic ketos isomerase (Rhodobacter pharaonis)]	63	51	132
3011	1	440	252	lgi 1480237	[phenylacetaidhyde dehydrogenase [Escherichia coli]]	63	40	169
3078	1	609	400	lgi 1487982	[intrinsic membrane protein [Nycphales hominis]]	63	36	210
3139	1	2	217	lgi 439126	[glutamate synthase (NADPH) [Acospirillum brasiliense piri A49916 A59916]] - <i>zospirillum brasiliense</i>	63	47	216
1625	1	791	198	lgi 623071	[lif 623071; initiation factor [Influenza A-11]]	63	48	196
3658	1	1	399	lgi 1305697	[lrrKA (Bacillus subtilis)]	63	37	399
3659	1	1	395	lgi 1256135	[lrbP (Bacillus subtilis)]	63	48	393
3781	1	220	161	lgi 1256902	[pyruvate decarboxylase isozyme 2 (Swiss Prot accession number P14467)]	63	36	360
					[Saccharomyces cerevisiae]			
3900	1	338	171	lgi P10537 ANB-	[BETA-ANYLASE (EC 3.2.1.2) (1,4-ALPHA-D-GLUCAN MALTOHYDROLASE)]	63	54	168
4309	1	3	176	lpir A37967 A379	[neural cell adhesion molecule Ng-CAM precursor - chicken]	63	57	174
4367	1	1	195	lgi 1121932	[larp gene product [Pichia pastoris]]	63	30	195
4422	1	1	312	lgi 151259	[lNG-CoA reductase (EC 1.1.1.81) (Pseudomonas nevlonii pI A4756 A4756 hydroxymethylglucaryl-CoA reductase (EC:1.1.1.81) Pseudomonas sp.)]	63	51	312
4468	1	6	108	lgi 29464	[ATPase (Lactococcus lactis)]	63	36	103
33	1	1411	2400	lgi 153675	[tagatose 6-P kinase (Streptococcus mutans)]	62	44	990
36	9	5985	6218	lgi 490521	[lrrKSA (lmono septins)]	62	51	234

TABLE 2

5
10
15
20
25
30
35
40
45
50
55

S. durans - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	t start	t ident	length (nt)
37	1	2	721	gi 1107531	[coccus product (Campylobacter coli)]	62	33	720
38	15	10912	11589	gi 1222058	[H. influenzae predicted coding region HIN1279 (Haemophilus influenzae)]	62	38	678
38	25	119526	120329	gi 695280	[ORF2 (Alcaligenes eutrophus)]	62	41	804
57	2	2523	1700	gi 171234	[orf1 (Haemophilus influenzae)]	62	55	744
57	9	6646	6350	gi 508174	[E1IB domain of PTS-dependent Gnt transport and phosphorylation Escherichia coli]	62	35	297
58	1	2	559	gi 755152	[highly hydrophobic integral membrane protein (Bacillus subtilis) sp P42551 macBACU TEICHOIC ACID TRANSLOCATION PENEASE PROTEIN AGC.]	62	34	558
67	10	8230	2014	gi 470683	[Shows similarity with ATP-binding proteins from other ABC-transporters, Swiss Prot Accession Numbers P2137, P04207, P04205, P24116 Escherichia coli]	62	34	765
69	8	8315	7454	gi 46816	[actVA 4 gene product (Streptomyces coelicolor)]	62	44	822
80	2	1793	1320	gi 3993	[LysP-N-acetyl muramoylalanine-D-glutamate ligase (Bacillus subtilis)]	62	43	474
87	7	7034	9205	gi 217191	[5'-nucleotidase precursor (Vibrio parahaemolyticus)]	62	18	2172
100	3	4051	3089	gi 1511047	[Lysophosphoglycerate dehydrogenase (Methanococcus jannaschii)]	62	42	963
102	1	2	520	gi 153555	[mismatch repair protein (Streptococcus pneumoniae plr C28667 C28667 DNA mismatch repair protein hexA - Streptococcus pneumoniae)]	62	34	519
112	2	466	1068	gi 153741	[ATP-binding protein (Streptococcus mutans)]	62	37	603
114	7	6055	7562	gi 120866	[L-fucose operon activator (Haemophilus influenzae)]	62	38	704
116	4	6823	5611	gi 677947	[ApdC (Bacillus subtilis)]	62	37	3191
124	8	6855	6004	gi 1653777	[product similar to E. coli PFAZ protein (Bacillus subtilis sp P4587 HEMK_BACSU) possible protoporphyrinogen oxidase (EC.1.3.3.1).]	62	44	852
148	1	24	554	gi 467156	[Unknown (Bacillus subtilis)]	62	30	531
149	20	7591	6725	gi 1203807	[replicative DNA helicase (Haemophilus influenzae)]	62	41	867
163	3	11503	1153	gi 40067	[X gene product (Bacillus subtilis)]	62	42	351
164	15	14673	15612	gi 42229	[P3 gene product (AA 1 - 314) (Escherichia coli)]	62	38	960
165	2	1166	1447	gi 403936	[phenylalanyl-tRNA synthetase alpha subunit (Gly294 variant) unidentified cloning vector]	62	38	282
166	2	2084	5089	gi 108861	[GRC start codon (Lactococcus lactis)]	62	44	1006
171	1	1225	614	gi 1016053	[hypothetical protein (SP P2049 Mycoplasma genitalium)]	62	41	612

TABLE 2

EP 0 786 519 A2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
183	1	2521	1310	gi 141045	hexY (Bacillus subtilis)	62	45	1212
200	1	3	1956	gi 142439	[ATP-dependent nuclelease (Bacillus subtilis)]	62	32	954
237	2	935	1966	gi 141695	[hIC protein (Escherichia coli)]	62	44	1032
261	3	4008	2605	gi 141121	[ORF A] putative (Bacillus firmus)	62	42	1404
299	8	4477	4719	gi 146741	[expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed (Bacillus subtilis) gi 146741] expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed Bacil	62	47	243
304	6	5018	3819	gi 153015	[FtsA protein (Staphylococcus aureus)]	62	43	1200
324	1	2	262	gi 142717	[cytochrome aa3 controlling protein (Bacillus subtilis) pir A3960 A3960] Cta protein - Bacillus subtilis sp P12946 CTRA_BACSU CYTOCHROME AA3 CONTROLLING PROTEIN.	62	30	261
325	2	269	1207	gi 158108	[methionyl-tRNA formyltransferase (Escherichia coli)]	62	39	919
332	6	4894	4631	gi 1499860	[uridine 5'-monophosphate synthase (Methanococcus jannaschii)]	62	36	264
355	1	2	170	gi 145925	[fcb (Escherichia coli)]	62	32	369
365	8	6628	6804	gi 1413943	[ipa-19d gene product (Bacillus subtilis)]	62	54	177
369	2	2744	1636	pir A43577 A35	[regulatory protein pfr - Clostridium perfringens]	62	42	1119
370	1	14	264	gi 140665	[beta-D-glucosidase (Clostridium thermocellum)]	62	37	231
415	3	2109	3176	gi 14205401	[transport ATP-binding protein (Haemophilus influenzae)]	62	35	468
429	1	1578	790	gi 1046024	[N-acetyltransferase subunit J (Mycoplasma genitalium)]	62	40	789
444	2	704	1369	gi 1561510	[nudotidyltransferase nod (Rhizobium loti)]	62	37	666
477	2	751	1869	pir A1840 A84	[ring-infected erythrocyte surface antigen 2. REKA-2 - <i>Plasmodium falciparum</i>]	62	44	1119
485	1	241	1707	gi 117934	[betaine aldehyde dehydrogenase (Beta vulgaris)]	62	43	1467
487	3	1141	1311	gi 149445	[ORF1 (Lactococcus lactis)]	62	31	171
494	2	1134	1313	gi 166815	[ribulose bisphosphate carboxylase/oxygenase activase (Arabidopsis thaliana)]	62	37	180
518	1	193	882	gi 153491	[O-methyltransferase (Streptomyces glaucescens)]	62	39	690
534	2	369	2522	gi 1480429	[putative transcriptional regulator (Bacillus stearothermophilus)]	62	35	2154
551	6	4371	4820	gi 151113	[ferric uptake regulation protein (Campylobacter jejuni)]	62	37	450
574	1	1	570	gi 153000	[lantibiotic B (Staphylococcus aureus)]	62	43	570

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins									
Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)	
590	1	2	134	1171	[gi 40367] ORFC [Clostridium acetobutylicum]	62	37	828	
655	1	196	810	[gi 147195]	[lpha protein (Escherichia coli)]	62	44	435	
656	1	2	478	[gi 1205451]	[cell division inhibitor (Listeria monocytogenes)]	62	36	477	
676	1	692	348	[gi 1511613]	[methyl coenzyme M reductase system component A2 (Methanococcus jannaschii)]	62	36	345	
687	1	493	248	[gi 49372]	[Aspartagine (Bacillus licheniformis)]	62	48	246	
700	1	267	944	[gi 1205822]	[hypothetical protein (GB:X75627-4) (Listeria monocytogenes)]	62	40	678	
840	2	1715	1041	[gi 105865]	[M genitalium predicted coding region NC1B1 (Mycoplasma genitalium)]	62	36	675	
864	4	898	1491	[gi 1144332]	[deoxyuridine nucleotidohydrolase (Homo sapiens)]	62	38	594	
916	1	35	400	[gi 141931]	[lpa-7d gene product (Bacillus subtilis)]	62	45	366	
1071	1	1	771	[gi 1510649]	[aspartokinase I (Methanococcus jannaschii)]	62	40	771	
1084	1	19	609	[gi 688011]	[ApN-1 antigen (human, infantile patient, testis, Peptida, 505 aa)]	62	39	591	
1103	1	3	203	[gi 581261]	[ORF homologous to E. coli mca (listeriolysin aurantiacus) P1r S14030 S14030]	62	51	201	
1217	1	463	233	[gi 1460025]	[lOHF2, putative (Streptococcus pneumoniae)]	62	41	231	
1533	1	644	414	[gi 1413968]	[lpa-4d gene product (Bacillus subtilis)]	62	48	211	
1537	1	1	257	[gi 1510641]	[laleny1-tRNA synthetase (Methanococcus jannaschii)]	62	29	255	
2287	1	3	161	[gi 45956]	[lmp1C gene product (Proteus mirabilis)]	62	45	159	
2386	1	3	245	[gi 125708]	[nontoxic component (Clostridium botulinum)]	62	31	243	
2484	1	331	167	[gi 120232]	[DNA-repair protein (recA) (Anabaena variabilis)]	62	35	165	
2490	1	798	400	[gi 515668]	[lpeB gene product (Staphylococcus epidermidis)]	62	42	399	
1016	1	596	300	[gi 110032]	[lucoperhydrolase III (Bacillus subtilis)]	62	51	297	
3116	1	1	213	[gi 466883]	[lfrS; Bl496_C2_193 (Mycobacterium leprae)]	62	44	213	
3297	1	823	413	[gi 475715]	[lacetyl coenzyme A acetyltransferase (thiolase) (Clostridium acetobutylicum)]	62	42	411	
1609	1	31	276	[gi 110801]	[homologous to N-acyl-L-amino acid amidohydrolase of Bacillus sphaericus]	62	48	246	
1665	2	584	402	[gi 151259]	[HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mesvalonil) P1r A4756 A4756]	62	40	183	
3733	1	3	374	[gi 125197]	[thioredoxin reductase (Listeria acidaminophilum)]	62	42	372	

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (ID)	Stop (ID)	match accession	match gene name	% sim	% ident	length (ntc)
3898	1	1	237	gi 153675	Uratase 6-P kinase [Streptococcus mutans]	62	45	237
4027	1	283	163	gi 30705	homologue to gene 30 (aa 1-59) / putative [Bovine herpesvirus 4]	62	43	141
4109	1	727	365	gi 41748	[hsdK protein (AA 1-520) Escherichia coli]	62	45	363
4303	1	1	303	gi 11301813	[Yqew Bacillus subtilis]	62	43	303
4380	1	530	267	gi 1215684	mevalonate pyrophosphate decarboxylase [Saccharomyces cerevisiae]	62	55	264
4494	1	2	256	gi 510692	[enterotoxin H Staphylococcus aureus]	62	34	235
4598	1	411	223	gi 765513	[ORF4: putative [Streptomyces violaceoruber]	62	45	189
4624	1	1	222	gi 41748	[hsd protein (AA 1-520) Escherichia coli]	62	45	222
5	5	4200	1912	gi 1920831	[ORF95: putative [lactococcus lactis phage B15-T1]	61	36	357
11	1	320	162	gi 613556 C333	[prochymosin alpha homolog (clone 32) - human (fragment)]	61	33	159
16	111	110291	113298	gi 11205391	[hypothetical protein (SP P31993) Haemophilus influenzae]	61	44	948
32	1	281	A01	gi 1106504	[exo-beta 1,3 glucanase (Cochliobolus carbonum)	61	50	519
38	3	616	1107	gi 11510864	[glutathione transport ATP-binding protein Q Methanococcus jannaschii]	61	41	492
45	4	3082	4038	gi 1109666	[PROX Bacillus subtilis]	61	45	957
48	6	7118	7504	gi 1498839	[ORF2 Clostridium perfringens]	61	33	387
51	9	4605	5570	gi 1304169	[trAC (Plasmid pAD1)]	61	42	956
60	6	1689	2243	gi 11205893	[hypothetical protein (DB U00011_31) Haemophilus influenzae]	61	32	555
62	9	5559	5122	gi 851656	[Na+/H+ antiporter system ORF2 (Bacillus alcalophilus)]	61	38	418
67	5	4330	5646	gi 1466612	[hika Escherichia coli]	61	36	1317
74	2	2400	1504	gi 1204816	[carbamate kinase [Haemophilus influenzae]	61	40	897
85	3	2198	1101	gi 1498756	[amidophosphotriesterase PurF Rhizobium etli]	61	41	1098
86	4	1995	1582	gi 11499931	[H. jannaschii predicted coding region MJ1083 Methanococcus jannaschii]	61	44	414
97	1	74	649	gi 1518679	[orf Bacillus subtilis]	61	44	576
99	2	2454	1990	gi 1413958	[lpa-3d gene product Bacillus subtilis]	61	18	465
124	7	6223	5123	gi 1568801	Similar to Saccharomyces cerevisiae SWS protein [Bacillus subtilis] gi 5491581 S19358 1pc-2D protein - Bacillus subtilis sp P31951 YMLC-BACSU HYPOTHETICAL 37.0 KD PROTEIN IN SPOUT-GLYC INTERGENIC REGION.	61	46	1101
125	4	1668	2521	gi 11491613	[ORF4 gene product Chordoflexus aurantiacus]	61	43	864

TABLE 2

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
132	1	1250	627	pit W00259 P002	[hypothetical protein 1 - Enterococcus faecalis plasmid pAM-beta-1 (fragment)]	61	43	624
149	9	1617	2075	gi 1143132	[deoxyuridine nucleotidohydrolase [Homo sapiens]]	61	40	543
149	22	8690	2869	gi 160047	[p101/acidic basic repeat antigen (Plasmodium falciparum) pif A29232 A29232 [strain Camp]]	61	35	822
166	3	1915	2361	gi 1439694	[10K malaria antigen precursor - Plasmodium vivax [strain 3D7]]	61	37	447
171	9	9675	17948	gi 1467446	[H11 protein, member of the H11T-family [Methanococcus jannaschii] similar to SpvB [Bacillus subtilis]]	61	38	1728
174	1	1042	2340	gi 1213374	[glutaryl 7-ACA acylase precursor [Bacillus laterosporus]]	61	49	1299
190	4	5034	4111	gi 1402886	[bmuU [Bacillus subtilis]]	61	37	924
216	1	2	120	gi 141861	[leukopyric initiation factor 2 beta [orf-2 beta] [Oryctolagus cuniculus]]	61	29	169
227	7	4161	5048	gi 1216341	[ORF for methionine amino peptidase [Bacillus subtilis]]	61	41	888
238	4	1959	3047	gi 140543	[cbic protein [Erwinia chrysanthemi]]	61	38	1089
247	1	2	694	gi 1537231	[ORF f579 [Escherichia coli]]	61	38	693
247	2	678	1034	gi 142226	[rho protein [Agrobacterium tumefaciens]]	61	40	357
257	2	1523	2627	gi 69379	[lgrv-1 protein [Hycobacterium leprae]]	61	40	897
268	2	3419	3051	gi 140664	[ORFAl (Clostridium acetobutylicum)]	61	41	369
275	4	4621	4827	gi 1124888	[hypothetical protein (G:487049.57) [Haemophilus influenzae]]	61	36	207
277	1	1	1H45	gi 7H4897	[beta-N-Acetylhexosaminidase [Streptococcus pneumoniae] p11 56190 A56190 [mannosyl-glycoprotein nbo-beta-N-acetylglucosaminidase [EC 3.2.1.96]] precursor - streptococcus pneumoniae]	61	45	1H45
278	19	8003	1032	gi 1467462	[leucine synthetase A [Bacillus subtilis]]	61	43	972
278	10	9878	9535	gi 1205919	[Na ⁺ and Cl ⁻ dependent gamma-aminobutyric acid transporter [Haemophilus influenzae]]	61	38	1344
283	1	1	366	gi 755607	[polyA polymerase [Bacillus subtilis]]	61	36	366
285	2	1918	1496	gi 386108	[cell wall enzyme - Enterococcus faecalis]	61	43	423
291	1	86	334	gi 145265	[PBP3 [Pectinia hybrida]]	61	38	249
318	1	1104	694	gi 290531	[similar to beta-glucoside transport protein [Escherichia coli] similar to beta-glucoside transport protein [Escherichia coli]]	61	47	411
330	2	1912	1190	gi 1001805	[solB 451 PBP ECOLI PTS SYSTEM, ARABININ-LIKE IIB COMPONENT PHOSPHOTRANSFERASE ENZYME II, B COMPONENT [EC 2.7.1.69]. hypothetical protein [Synchocystis sp.]	61	41	723

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	start	ident	length (nt)
10								
185	2	1513	1025	[gi 531098	[DnaD protein (Bacillus subtilis)]	61	42	489
				[gi 1103853	[Tqff (Bacillus subtilis)]	61	44	296
426	1	794	399	[gi 1103853	[Tqff (Bacillus subtilis)]	61	36	612
				[gi 73522	[Inositol-4,5-diphosphate 1-kinase (Dictyostelium discoideum)]	61	30	789
464	2	784	560	[gi 1121120	[CSJ87.5 gene product (Caenorhabditis elegans)]	61	387	225
				[gi 6077	[ORF60; putative (Bacteriophage LL-HI)]	61	47	1281
470	8	6077	7357	[gi 613073	[Unknown (Bacillus subtilis)]	61	45	276
				[gi 467484	[anthranilate synthase glutamine amidotransferase (Acinetobacter alcoaceticus)]	61	42	621
509	1	554	279	[gi 1103800	[Unknown (Bacillus subtilis)]	61	33	159
				[gi 1103800	[Unknown (Bacillus subtilis)]	61	32	271
555	3	1916	1296	[gi 1103800	[Unknown (Bacillus subtilis)]	61	33	315
				[gi 1711	[R225_C2_195 (Mycobacterium leprae)]	61	47	655
569	1	857	857	[gi 467090	[SURVIVAL PROTEIN SURF MONOLOG (FRAGMENT)]	61	37	1233
				[sp P16666 S0RE-	[Immunity repressor protein (Haemophilus influenzae)]	61	37	1233
585	2	961	803	[sp P16666 S0RE-	[Unknown (Bacillus subtilis)]	61	33	159
				[gi 11211602	[Unknown (Bacillus subtilis)]	61	32	271
592	3	1694	1432	[gi 11211602	[Unknown (Bacillus subtilis)]	61	33	315
				[gi 1103843	[Unknown (Bacillus subtilis)]	61	37	1233
603	1	43	357	[gi 1103718	[Unknown (Vibrio parahaemolyticus)]	61	33	159
				[gi 1146243	[Unknown (Bacillus subtilis)]	61	37	1233
669	1	2467	1235	[gi 1146243	[Unknown (Bacillus subtilis)]	61	33	159
				[gi 1103373	[Unknown (Bacillus subtilis)]	61	36	297
675	3	805	1101	[gi 1103373	[Unknown (Bacillus subtilis)]	61	36	297
				[gi 537251 S37251	[Unknown (Bacillus subtilis)]	61	32	626
703	1	1656	829	[gi 537181	[ORF_f470 (Escherichia coli)]	61	32	626
				[gi 806881	[DNA polymerase I (Bacillus stearothermophilus)]	61	39	613
728	1	1628	816	[gi 806881	[DNA polymerase I (Bacillus stearothermophilus)]	61	38	540
				[gi 709992	[Hypothetical protein (Bacillus subtilis)]	61	26	312
821	1	61	318	[gi 709992	[Hypothetical protein (Bacillus subtilis)]	61	18	258
				[gi 606397	[Portal protein gp3 (Bacteriophage HK97)]	61	40	747
856	2	2313	1567	[gi 606397	[Unknown (Mycoplasma genitalium)]	61	32	271
				[gi 1103213	[Unknown (Bacillus subtilis)]	61	38	540
923	1	1081	542	[gi 1103213	[Unknown (Bacillus subtilis)]	61	34	273
				[gi 1107541	[CJ309.8 (Caenorhabditis elegans)]	61	31	288
1124	1	59	370	[gi 1107541	[Unknown (Bacillus subtilis)]	61	55	897
				[gi 1103826	[Unknown (Staphylococcus aureus)]	61	41	246
1492	1	548	276	[gi 1103826	[Unknown (Staphylococcus aureus)]	61	32	271
				[gi 73522	[Inositol-4,5-diphosphate 1-kinase (Dictyostelium discoideum)]	61	34	273
1602	1	46	318	[gi 73522	[Unknown (Bacillus subtilis)]	61	31	271
				[gi 1045964	[Hypothetical protein (GB:U14003;297) (Mycoplasma genitalium)]	61	31	288
2500	1	577	290	[gi 1045964	[Unknown (Bacillus subtilis)]	61	31	288
				[gi 1103826	[Unknown (Staphylococcus aureus)]	61	32	271
3076	1	3	248	[gi 1103826	[Unknown (Lactococcus lactis)]	61	32	271
				[gi 1103826	[Unknown (Lactococcus lactis)]	61	32	271

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	Batch accession	Batch gene name	Ident	length (nt)
1609	2	207	401	gi 1408501	homologous to N-acyl-L-amino acid amidohydrolase of <i>Bacillus stearothermophilus</i> [Bacillus subtilis]	61	39
3662	1	1477	710	gi 130813	[Qew] [Bacillus subtilis]	61	195
3672	1	2	442	gi 78897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir AS390 AS6390 mannosyl-glycoprotein nido-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	42
3724	1	2	220	gi 1109366	respiratory nitrate reductase [Bacillus subtilis]	61	738
3728	1	3	398	gi 167943	AppB [Bacillus subtilis]	61	441
3884	1	3	401	gi 78897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir AS390 AS6390 mannosyl-glycoprotein nido-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	396
3971	1	3	383	gi 78897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir AS390 AS6390 mannosyl-glycoprotein nido-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	47
4038	1	661	359	gi 1133997n	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	61	199
4041	1	546	274	gi 1413953	lmu-23d gene product [Bacillus subtilis]	61	219
4047	1	1	402	gi 1529991	Unknown [Bacillus subtilis]	61	46
4102	1	1	345	gi 1970225	rrsA [Escherichia coli]	61	402
4155	1	1	336	gi 78897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir AS390 AS6390 mannosyl-glycoprotein nido-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	345
4188	1	463	233	gi 1456688	hsa gene of Ecoprt gene product [Escherichia coli] pir S3837 S38437 hidM precursor - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli (SUS 40-520)	61	336
4374	1	542	273	gi 78897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir AS390 AS6390 mannosyl-glycoprotein nido-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	270
4389	1	2	172	gi 147516	ribokinase [Escherichia coli]	61	231
4621	1	2	268	gi 78897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir AS390 AS6390 mannosyl-glycoprotein nido-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	171
4663	1	27	227	gi 1976075	hsa [Escherichia coli]	61	267
4	6	6663	5536	gi 1408501	homologous to N-acyl-L-amino acid amidohydrolase of <i>Bacillus subtilis</i> [Bacillus subtilis]	60	50
						60	201
						60	1126

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
11	6	3426	3725	gi 410748	ring-infected erythrocyte surface antigen (Plasmodium falciparum) (alias) ring-infected erythrocyte surface antigen (strain PC27/Papua New Guinea) sp P13830 RESA_PLAFF	60	24	300
11	14	11035	10313	gi 11217651	Plasmodium falciparum (strain PC27/Papua New Guinea) sp P13830 RESA_PLAFF	60	28	723
16	12	11917	12920	gi 1001453	carbonyl reductase (NADPH) (Rattus norvegicus)	60	37	1014
13	1	26	469	gi 1386109	hypothetical protein (Symbiocystis sp.)	60	41	444
37	13	10414	9834	gi 11336556	carbonyl reductase (NADPH) (Rattus norvegicus)	60	40	981
39	4	4364	4522	gi 14872	hypothetical protein (Saccharomyces kuyveri)	60	47	159
41	1	2047	1025	gi 1112822	regulatory protein (Enterococcus faecalis)	60	39	1023
43	4	2474	3607	gi 148016	D-alanine racemase cds (Bacillus subtilis)	60	40	1134
44	10	6756	7769	gi 1414234	D-alanine racemase esterase (Bacillus subtilis)	60	52	1014
45	10	8874	9074	gi 1303949	D-alanine racemase esterase (Bacillus subtilis)	60	44	201
56	18	127842	124430	gi 148764	D-alanine racemase product (Rhizobium meliloti)	60	35	1413
60	2	173	388	gi 1130864	D-alanine racemase (Bacillus subtilis)	60	33	216
61	2	1357	1619	gi 1467124	D-alanine racemase (Bacillus subtilis)	60	43	1263
69	1	787	395	gi 11518853	D-alanine lythimurum	60	36	193
NA	1	1	1188	gi 11480429	putative transcriptional regulator (Bacillus stearothermophilus)	60	30	1168
92	1	6	4735	gi 1349227	transmembrane protein (Escherichia coli)	60	37	855
92	1	7	5936	gi 1466613	trkA (Escherichia coli)	60	38	1014
93	1	1	949	gi 11510925	triosephosphate isomerase (Bacillus subtilis)	60	27	474
96	1	6	7166	gi 171715	triosephosphate isomerase (Klebsiella pneumoniae)	60	30	213
98	1	6	3212	gi 1467425	triosephosphate isomerase (Bacillus subtilis)	60	42	858
102	10	7158	7430	gi 143092	triosephosphate isomerase small subunit (Bacillus subtilis)	60	37	271
109	11	9127	10515	gi 11253259	triosephosphate isomerase small subunit (EC 5.3.1.8) (alias) tricetylhydroxy-acid synthase small subunit (EC 5.3.1.8)	60	28	1369
109	12	10499	11656	gi 1141954	triosephosphate isomerase (Staphylococcus aureus)	60	41	1158
119	2	4630	3134	gi 11524260	triosephosphate isomerase (Mycobacterium tuberculosis)	60	45	1497

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
121	9	6957	7046	[gi 1107539	[cauC gene product (Campylobacter coli)]	60	35	690
140	7	7704	6013	[gi 146547	[kdpA (Escherichia coli)]	60	45	1692
145	1	2	703	[gi 1460077	[unknown (Mycobacterium tuberculosis)]	60	23	702
150	3	2809	2216	[gi 1146220	[putative (Bacillus subtilis)]	60	40	594
157	2	1389	961	[gi 1303975	[YQ1X (Bacillus subtilis)]	60	30	429
158	5	5125	4769	[gi 1449288	[unknown (Mycobacterium tuberculosis)]	60	36	357
159	1	511	257	[gi 580932	[mucD gene product (Bacillus subtilis)]	60	43	255
160	1	159	1187	[gi 1204532	[hypothetical protein (GB:Li9201_29) (Haemophilus influenzae)]	60	34	1029
161	14	8249	7866	[gi 1896003	[ORF3: Putative oligopeptidase based on homology with Lactococcus lactis POF (GenBank Accession Number J32522) (caldicelluloseruptor saccharolyticus)]	60	34	384
172	1	1331	2110	[gi 148280	[28.2 kDa protein (Streptococcus pneumoniae)]	60	33	780
173	2	4082	2460	[gi 1524397	[glycine betaine transporter OpdU (Bacillus subtilis)]	60	41	1623
173	4	5963	4953	[gi 1100777	[NAUP dependent leukotene B ₄ 12-hydroxydehydrogenase (Sus scrofa)]	60	44	1011
198	1	3	995	[gi 413943	[lpa-15d gene product (Bacillus subtilis)]	60	42	993
201	4	3641	4573	[sp P27028 YADT_	[HYPOTHETICAL 29.4 KD PROTEIN IN HEMI-PFS INTERGENIC REGION PRECURSOR.]	60	37	933
203	1	3269	2415	[gi 1227798	[CD729_34p_CAI_0_14 (Serratia cerevalis)]	60	43	875
206	9	12234	12515	[sp P7347 YEDC_	[HYPOTHETICAL 21. KD PROTEIN IN ASPS 5 REGION]	60	47	282
212	4	1213	1410	[gi 1327211	[hemagglutinin-neuramidase fusion protein (Human paramyxovirus 3)]	60	34	198
214	1	65	1153	[gi 1204366	[hypothetical protein (GB:U14003_130) (Haemophilus influenzae)]	60	36	1089
237	1	2	937	[gi 1493377	[H15D (Lactococcus lactis)]	60	40	936
241	6	5696	4998	[gi 1046160	[hypothetical protein (GB:U00021_5) (Mycoplana genitalium)]	60	37	699
260	6	5919	6485	[gi 431950	[similar to a B. subtilis gene (GB: BACHEMENY_5) (Clostridium acetivoranum)]	60	35	567
264	1	2432	1218	[gi 397526	[clumping factor (Staphylococcus aureus)]	60	53	1215
267	1	3	1409	[gi 148316	[Na ⁺ -antporter protein (Enterococcus hirae)]	60	27	1407
275	3	3804	4595	[pif F1689 F368	[leuD 3'-region hypothetical protein - Lactococcus lactis subsp. lactis (strain 11401)]	60	35	792
291	3	860	1198	[gi 1208689	[cod1 for C. elegans cDNA y130e12.3; contains CH2-type zinc fingers (Caenorhabditis elegans)]	60	33	119

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (Int)	Stop (Int)	match accession	match gene name	% sim	% ident	length (nt)
307	6	3421	3176	[gi 11070014	protein-dependent [Bacillus subtilis]	60	36	246
316	8	4957	5823	[gi 413952	lipo-2Bd gene product [Bacillus subtilis]	60	41	867
326	4	2996	3484	[gi 11204464	membrane-associated component, branched amino acid transport system [Haemophilus influenzae]	60	39	489
332	5	4887	4363	[gi 11205449	l colicin V production protein [pur regulon] [Haemophilus influenzae]	60	37	525
357	1	1062	532	[gi 11081842	single-stranded DNA specific exonuclease [Escherichia coli]	60	41	531
375	2	96	362	[gi 110837	adenylyl cyclase gene product [Saccharomyces kluyveri] r [Dolittle]	60	47	267
397	1	66	416	[gi 1109999	adenylate cyclase [EC 4.6.1.1] - yeast cytochrome kluyveri]	60	37	351
409	1	2	163	[gi 1109700	Glucarate dehydratase [Bacillus subtilis]	60	35	162
453	4	914	1237	[gi 11196899	unknown protein [Staphylococcus aureus]	60	36	324
453	7	3838	3620	[sp P12222 CFL]	HYPOTHETICAL 226 KD PROTEIN (ORF 1901)	60	31	219
470	1	622	945	[p11 S0782 S307	integrin homolog - yeast [Saccharomyces cerevisiae]	60	31	324
500	1	118	606	[gi 1107407	unknown [Bacillus subtilis]	60	36	469
503	3	752	982	[gi 1107835	lysozyme heavy chain [Dictyostelium discoideum]	60	34	231
505	4	2238	3563	[gi 11150732	NADH oxidase [Methanococcus jannaschii]	60	26	1326
523	1	3	1043	[gi 1103331	alkaline phosphatase regulatory protein phoR - <i>Bacillus subtilis</i> sp P23545 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN NOR (EC 2.7.3.1).	60	41	1041
543	1	1	465	[gi 1111103	cobalt transport ATP-binding protein o [Methanococcus jannaschii]	60	40	465
545	1	1	726	[gi 11098192	putative [Pseudomonas aeruginosa]	60	40	726
556	1	2	1054	[gi 1107402	tax gene product [Bordetella pertussis]	60	42	1053
578	1	574	489	[gi 11025129	H. influenzae predicted coding region H10882 [Haemophilus influenzae]	60	42	486
594	1	1	624	[gi 11212755	adenylyl cyclase [Neurospora hydrophila]	60	45	624
604	1	3	530	[gi 1145925	[fecB [Escherichia coli]	60	42	528
620	1	926	465	[gi 11025483	bicyclomycin resistance protein [Haemophilus influenzae]	60	33	462
630	2	871	1122	[gi 11086242	unknown [Bacillus subtilis]	60	41	252
645	2	574	425	[gi 1105336	lysine hydroxymethyltransferase (lysine methylase) [Haemophilus influenzae]	60	28	150

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% Ident	length (nt)
10								
684	1	1082	843	gi 1205538	hypothetical protein [Bacillus subtilis]	60	19	240
					orf111 gene product [Bacillus subtilis]	60	46	483
786	1	967	405	gi 1402944	hypothetical protein [Bacillus subtilis]	60	40	243
844	1	568	346	gi 790843	urea amidolyase [Bacillus subtilis]	60	41	726
851	1	1	726	gi 1528661	[GDP reductase [Ascaris lumbricoides]	60	39	871
871	1	1746	874	gi 11001493	hypothetical protein [Synechocystis sp.]	60	39	871
896	1	1558	839	gi 104926	[NADH dehydrogenase, subunit 5 [Schizophyllum commune] sp P50368 NUSH_SCRCO	60	39	720
					NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 [EC : 6.5.3.1]	60	30	554
908	2	446	753	gi 663880	novel hemolytic factor [Bacillus cereus]	60	31	306
979	1	2	595	gi 1429255	[putative orf1 [Bacillus subtilis]	60	40	168
					inner membrane copper tolerance protein [Escherichia coli] gi 871029	60	40	168
1078	1	669	502	gi 1581055	inner membrane copper tolerance protein [Escherichia coli] pir S47295 S47295	60	39	377
					inner membrane copper tolerance protein - escherichia coli	60	39	377
1112	1	1150	620	gi 107885	[ORF1 [Streptomyces griseus]	60	34	531
					[putative orf1 [Streptomyces griseus]	60	36	210
1135	1	484	275	gi 1171407	[putative isochorotidate cyclase]	60	36	546
1146	1	17	562	gi 1239981	hypothetical protein [Bacillus subtilis]	60	30	357
1291	1	716	360	gi 1557530 S575	[carboxyl esterase - Actinobacter calcoaceticus	60	44	168
1332	1	336	169	gi 1222056	[fumonitrilase [Haemophilus influenzae]	60	39	144
1429	1	3	146	gi 1205619	[ferritin like protein [Haemophilus influenzae]	60	36	285
1722	1	570	286	gi 240052	[dihydroflavonol-4-reductase, DFR [Hordeum vulgare-barley, cv. Gale, sp. id. 354 aa]	60	36	285
					ORF 1. [Plasmid pQ01]	60	20	186
2350	1	385	1200	gi 1971226	[prophenate dehydratase [Bacillus subtilis]	60	48	210
2916	1	519	310	gi 1508681	[putative orf1 [Bacillus subtilis]	60	37	267
3027	1	568	302	gi 1146199	[putative orf1 [Bacillus subtilis]	60	51	189
3084	1	20	1208	gi 1407784	[orf1: novel antigen [Staphylococcus aureus]	60	34	225
					[cytadherence-accessory protein [Mycoplasma genitalium]	60	42	183
3155	1	2	226	gi 1046097	[mitochondrial long-chain enoyl-CoA hydratase/-hydroxacyl-CoA hydratase]	60	42	243
3603	1	368	186	gi 510108	[alpha-subunit [Rattus norvegicus]	60	36	144
					[IMG-CoA reductase [EC 1.1.1.68] (Pseudomonas methanolic) pir A4756 A44756]	60	36	144
3665	1	486	244	gi 151259	[hydroxymethylglutaryl-CoA reductase (EC 1.1.1.68) Pseudomonas sp]	60	36	144
3747	1	3	146	gi 174192	[lucc gene product [Escherichia coli]	60	36	144

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	orf	Start	Stop	match	match gene name	% sim	% ident	length
orf	orf	(nt)	(nt)	junction		(nt)	(nt)	(nt)
ID	ID							
3912	1	1	3	gi 1488695	novel antigen; orf-2 [Staphylococcus aureus]	60	44	333
4072	1	1	3	272	gi 1405879 [Escherichia coli]	60	33	270
4134	1	510	352	gi 180656	chemoreceptor protein [Rhizobium leguminosarum bv. viciae]	60	28	159
4207	2	677	402	gi 162031	similar to trimethylamine DH [Mycoplasma capricolum] pir S4995 S49950 probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolum (SGC1) (fragment)	60	41	276
4243	1	127	324	gi 1893317	peptide synthetase module [Micrococcus aeruginosae] pir S4911 S4911 probable amino acid activating domain - lycopodium aeruginosae (fragment) (SUB 144-528)	60	42	198
4110	1	624	313	gi 150980	lphB [Bacillus subtilis]	60	28	312
4145	1	341	173	gi 1510108	mitochondrial long-chain enoyl-CoA hydratase/-hydroxyl-CoA hydratase alpha-subunit [Rattus norvegicus]	60	42	171
4182	1	498	280	gi 17382	lacyl-CoA-dehydrogenase [Streptomyces purpurascens]	60	48	219
4474	1	53	223	gi 1510108	mitochondrial long-chain enoyl-CoA hydratase/-hydroxyl-CoA hydratase alpha-subunit [Rattus norvegicus]	60	42	171
23	4	4518	1523	gi 1426446	VIPB protein [Salmonella typhi]	59	39	996
33	2	707	1483	pir S4804 S486	hypothetical protein - Mycoplasma capricolum (SGC3) (fragment)	59	33	777
33	5	4651	5853	gi 16721	FS9B2.3 [Ceanorhabditis elegans]	59	33	1203
17	2	3326	2299	gi 142433	ORF2 [Bacillus subtilis]	59	37	910
38	121	16784	16593	gi 1912576	BIP [Phaeodactylum tricornutum]	59	40	-
52	1	3	2648	gi 1536972	ORF_090 [Escherichia coli]	59	44	100
54	122	14181	13402	gi 1403940	transcription regulator [Bacillus subtilis]	59	37	780
57	3	4397	3339	gi 1508176	Gat-1-P-DH, NAD dependent [Escherichia coli]	59	40	1059
66	1	986	495	gi 1103001	YgbT [Bacillus subtilis]	59	34	492
67	7	6552	7460	gi 1912461	lntK [Escherichia coli]	59	37	909
70	7	5383	6366	gi 1599822	PhD precursor [Rhizobium meliloti]	59	46	984
78	1	1	1449	gi 1971345	unknown, similar to <i>E. coli</i> cardiolipin synthase [Bacillus subtilis] NP P45160 YNE-BACSU HYPOTHETICAL 'S8.2' PROTEIN IN NAR1-ACDA NTERGENIC REGION	59	39	1449
82	110	11329	115334	gi 140338	lORF F (unidentified)	59	44	1206

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
89	2	1602	1958	[gi 642801]	unknown [<i>Saccharomyces cerevisiae</i>]	59	32	645
96	4	4940	5473	[gi 1333002]	protein of unknown function [<i>Rhodobacter capsulatus</i>]	59	33	534
98	1	2	820	[gi 1467421]	similar to <i>B. subtilis</i> DnaH [<i>Bacillus subtilis</i>]	59	34	819
119	1	166	1557	[gi 143122]	ORF B; putative [<i>Bacillus firmus</i>]	59	36	1392
120	10	6214	6756	[gi 15354]	[ORF 55.9 [<i>Bacteriophage T4</i>]]	59	39	543
120	16	112476	11510	[gi 1086575]	[Bata [<i>Rhizobium meliloti</i>]]	59	44	1035
123	1	386	195	[gi 984737]	[catalase [<i>Campylobacter jejuni</i>]]	59	38	192
130	1	370	645	[gi 1256634]	[23.8% identity over 120 aa with the <i>Synechococcus</i> sp. Np6v protein; putative [<i>Bacillus subtilis</i>]]	59	31	276
131	4	5278	5712	[gi 1510555]	[hypothetical protein (SP:PA2397) [<i>Methanococcus jannaschii</i>]]	59	39	435
164	1	3	509	[gi 1001342]	[hypothetical protein [<i>Synechocystis</i> sp.]]	59	41	507
164	4	1529	2821	[gi 1205165]	[hypothetical protein (SP:PA37764) [<i>Hanophillus influenzae</i>]]	59	35	1293
164	19	19643	21776	[gi 1001381]	[hypothetical protein [<i>Synechocystis</i> sp.]]	59	34	1714
173	3	4727	3717	[gi 1184121]	[auxin-induced protein [<i>Vigna radiata</i>]]	59	50	1011
179	2	2218	1668	[gi 143016]	[unidentified gene product [<i>Bacillus subtilis</i>]]	59	33	531
195	12	12669	11503	[gi 62778]	[NifS gene product [<i>Anabaena azollae</i>]]	59	41	1167
201	5	4702	5670	[gi 1510240]	[hemin permease [<i>Methanococcus jannaschii</i>]]	59	32	969
201	7	5719	6335	[gi 1511456]	[H. jannaschii predicted coding region H01437 [<i>Methanococcus jannaschii</i>]]	59	34	597
209	1	102	461	[gi 1201666]	[hypothetical protein (CA:K73124-52) [<i>Haemophilus influenzae</i>]]	59	42	360
214	3	1050	2234	[gi 1515133]	[2-nitropropane dioxygenase [<i>Willowsia saturnus</i>]]	59	36	1185
214	5	3391	4135	[gi 1103709]	[YrkJ [<i>Bacillus subtilis</i>]]	59	32	843
217	2	3381	2167	[gi 1290489]	[dtp (CC Site No. 18430) [<i>Escherichia coli</i>]]	59	44	1215
237	5	3078	3785	[gi 1491882]	[ltsA [<i>Lactococcus lactis</i>]]	59	38	708
251	2	376	960	[gi 1303791]	[YqjJ [<i>Bacillus subtilis</i>]]	59	34	585
286	1	1621	812	[gi 146551]	[transmembrane protein (kdpD) [<i>Escherichia coli</i>]]	59	31	810
316	5	4978	3860	[gi 105679]	[yeiH [<i>Escherichia coli</i>]]	59	32	1119
370	3	600	761	[gi 1303794]	[YqeM [<i>Bacillus subtilis</i>]]	59	35	162

TABLE 2

EP 0 786 519 A2

5. *autum* - putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
382	1	1009	506	gi 547513	orf3 [Haemophilus influenzae]	59	34	504
391	3	1620	1273	gi 1152901	[ORF 3] [Sphingobacter austrosibir]	59	37	348
406	3	2805	1705	gi 1709992	hypothetical protein [Bacillus subtilis]	59	34	1101
426	5	3802	3245	gi 1204610	iron(III) dicitrato transport ATP-binding protein FCE [Haemophilus influenzae]	59	36	558
429	2	1513	1148	gi 106809	homologous to sp NTRB_ECOLI [Escherichia coli]	59	42	366
440	2	708	1201	gi 466882	ppai: B149_C2_189 [Mycobacterium leprae]	59	37	594
461	4	2212	3135	gi 149295	homoserine kinase homolog [Streptococcus pneumoniae]	59	37	924
473	1	2929	1607	gi 147989	trigger factor [Escherichia coli]	59	40	1323
480	8	5882	6110	gi 1205311	[3R] -hydroxymyristol acyl carrier protein dehydrogenase [Haemophilus influenzae]	59	40	249
521	1	14	1354	gi 125620 gi 2456	staphylococcal - Staphylococcus aureus (fragment)	59	22	1341
534	4	2394	4073	gi 152746	mannitol-phosphate dehydrogenase [Streptococcus mutans] pir C44798 C44798	59	36	1080
535	1	1	954	gi 1469939	group B oligopeptidase PapB [Streptococcus agalactiae]	59	33	954
551	3	2836	3186	gi 1204511	bacterioferritin comigratory protein [Haemophilus influenzae]	59	45	351
573	2	449	940	gi 186681	[ORF YAL022] [Saccharomyces cerevisiae]	59	36	492
650	1	5	748	gi 1396400	similar to eukaryotic Na ⁺ /H ⁺ exchanger [Escherichia coli] sp P1270 YCEC_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SDR-ACS INTERGENIC REGION (0494).	59	10	744
664	1	566	285	gi 1262748	[lukF-PV like component [Staphylococcus aureus]	59	33	282
670	1	3	455	gi 122758	lukH [Bacillus subtilis]	59	42	453
674	3	543	929	gi 1293033	integrase [Bacteriophage phi-X174]	59	46	387
758	1	349	176	gi 1500472	[H. jannaschii] predicted coding region KJ1577 [Methanococcus jannaschii]	59	37	174
771	2	2270	1461	gi 522150	bromoperoxidase BPO-A1 [Streptomyces aureofaciens] sp P133912 BPA1_STRAU	59	44	810
					[NON-HAEM BROMOPEROXIDASE BRO-A1 (EC 1.11.1.-) BROMIDE PEROXIDASE] (BPO1). (SUB 2-275)			
825	1	2191	1097	gi 197526	[clumping factor [Staphylococcus aureus]	59	47	1095
1052	2	1094	723	gi 1289262	[cone ORF] [Bacillus subtilis]	59	36	372
1152	1	373	188	gi 1276668	[ORF238 gene product [Porphyra purpurea]	59	37	186

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start ID	Stop ID	Start (nt)	Stop (nt)	match accession	match gene name	Ident	length (nt)	
1198	1	492	247	gi 142439	gi 1043942	[ATP-dependent nuclelease (Bacillus subtilis)]	[tracyl-glycerol lipase (Galactomyces geotrichum)]	59	26	
1441	1	468	235	gi 142439	gi 1043942	[tracyl-tRNA synthetase (Mycobacteria genitalium)]	[tracyl-glycerol lipase (Galactomyces geotrichum)]	59	246	
2101	1	1	186	gi 1459250	gi 1043942	[tracyl-tRNA synthetase (Mycobacteria genitalium)]	[tracyl-glycerol lipase (Galactomyces geotrichum)]	59	17	
2205	1	793	398	gi 1303794	gi 1043942	[tracyl-tRNA synthetase (Mycobacteria genitalium)]	[tracyl-glycerol lipase (Galactomyces geotrichum)]	59	186	
2578	1	484	284	gi 258003	gi 1258003	[insulin-like growth factor binding protein complex acid-labile subunit (rats, liver, Peptide, 603 aa)]	[insulin-like growth factor binding protein complex acid-labile subunit (rats, liver, Peptide, 603 aa)]	59	396	
2967	1	2	145	340	gi 1212730	gi 1212730	[YqhK (Bacillus subtilis)]	[YqhK (Bacillus subtilis)]	59	48
3012	1	1	3	248	gi 173371	gi 173371	[neurofilament protein NF70 (Helix aspersa)]	[neurofilament protein NF70 (Helix aspersa)]	59	44
3544	1	1	3	401	gi 105218	gi 105218	[crotonase (Clostridium acetobutylicum)]	[crotonase (Clostridium acetobutylicum)]	59	31
3548	1	1	3	401	gi 105218	gi 105218	[crotonase (Clostridium acetobutylicum)]	[crotonase (Clostridium acetobutylicum)]	59	246
3580	1	698	351	gi 105218	gi 105218	[crotonase (Clostridium acetobutylicum)]	[crotonase (Clostridium acetobutylicum)]	59	199	
3720	1	1	722	363	gi 140494	gi 140494	[homologous to penicillin acylase (Bacillus subtilis)]	[homologous to penicillin acylase (Bacillus subtilis)]	59	42
4171	1	1	3	296	gi 105218	gi 105218	[crotonase (Clostridium acetobutylicum)]	[crotonase (Clostridium acetobutylicum)]	59	39
4305	1	1	618	310	gi 152193	gi 152193	[unknown (Mycobacterium tuberculosis)]	[unknown (Mycobacterium tuberculosis)]	59	42
18	1	1242	622	gi 145913	gi 145913	[N-acetylglucosamine transport protein (Escherichia coli) B (EC 7.1.1.69, N-acetylglucosamine-specific translocase, EC 9.99.23.1PTAA, E. coli PTS system, N-acetylglucosamine-specific IIABC component (E. coli))]	[N-acetylglucosamine transport protein (Escherichia coli) B (EC 7.1.1.69, N-acetylglucosamine-specific translocase, EC 9.99.23.1PTAA, E. coli PTS system, N-acetylglucosamine-specific IIABC component (E. coli))]	59	294	
20	1	7	7020	5815	gi 50502	gi 50502	[collagen alpha chain precursor IAA-27 (in 11271 (Mus musculus))]	[collagen alpha chain precursor IAA-27 (in 11271 (Mus musculus))]	59	39
21	5	3234	3636	gi 1054860	gi 1054860	[liphosphoribosyl anthranilate isomerase (Thermotoga maritima)]	[liphosphoribosyl anthranilate isomerase (Thermotoga maritima)]	59	309	
23	1	2	2041	11659	gi 11270880	gi 11270880	[EsgC (Streptococcus thermophilus)]	[EsgC (Streptococcus thermophilus)]	59	1176
23	10	1	9301	8090	[pir A1113]A111	[pir A1113]A111	[diaminopimelate decarboxylase (EC 4.1.1.20) - Pseudomonas aeruginosa]	[diaminopimelate decarboxylase (EC 4.1.1.20) - Pseudomonas aeruginosa]	58	43
3A	29	122555	122644	gi 973349	gi 973349	[vestitone reductase (Medicago sativa)]	[vestitone reductase (Medicago sativa)]	58	621	
44	1	1	2	406	gi 289727	gi 289727	[tertIcrone-binding protein (Bacillus subtilis)]	[tertIcrone-binding protein (Bacillus subtilis)]	58	37
45	1	1	552	gi 159444	gi 159444	[embryonic myosin heavy chain (1085 AA) (homo sapiens) 1x [S12460 S12460]]	[embryonic myosin heavy chain (1085 AA) (homo sapiens) 1x [S12460 S12460]]	58	33	
55	1	2	759	538	gi 158552	gi 158552	[galactose regulated protein (Echinococcus multilocularis)]	[galactose regulated protein (Echinococcus multilocularis)]	58	552
62	13	1	893	8068	gi 197553	gi 197553	[kinase-associated protein B (Bacillus subtilis)]	[kinase-associated protein B (Bacillus subtilis)]	58	32
63	3	1553	1717	gi 166326	gi 166326	[Arabidopsis thaliana unidentified mRNA sequence, complete cds I, one product (Arabidopsis thaliana)]	[Arabidopsis thaliana unidentified mRNA sequence, complete cds I, one product (Arabidopsis thaliana)]	58	222	
									165	

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (In)	Stop (In)	match accession	match gene name	% sim	% ident	length (nt)
67	13	112017	111229	gi 1228083	[NADH dehydrogenase subunit 2 (Chorthippus parallelus)]	58	41	749
96	8	18208	9167	gi 70992	[hypothetical protein <i>Bacillus subtilis</i>]	58	42	960
107	2	2065	13564	gi 806127	[<i>Escherichia coli</i> hprA gene for a protein similar to yeast PR16 and KP22]	58	37	702
112	7	4519	5613	gi 155588	[glucose-fructose oxidoreductase (<i>Zymomonas mobilis</i> pvtA)] [glucose-fructose oxidoreductase (EC 1.1.1.1) precursor - <i>Zymomonas mobilis</i>]	58	38	1095
114	6	7118	6503	gi 1377843	[<i>Escherichia coli</i> hprA]	58	38	816
143	2	2261	1395	gi 145605 A456	[mature-parasite-infected erythrocyte surface antigen HESA - <i>Plasmodium falciparum</i>]	58	31	867
151	2	717	950	gi 1370261	[<i>Unknown</i> (<i>Mycobacterium tuberculosis</i>)]	58	31	224
154	6	6015	4827	gi 11209227	[pCH101 gene product (<i>Chlamydia trachomatis</i>)]	58	41	1369
154	16	114281	133541	gi 1146613	[DNA ligase (EC 6.5.1.2) (<i>Escherichia coli</i>)]	58	39	741
155	3	2269	1892	gi 1303937	[YQ18 (<i>Bacillus subtilis</i>)]	58	34	378
174	1	1056	529	gi 190198	[hypothetical protein (<i>Bacillus subtilis</i>)]	58	26	528
189	4	1533	1769	gi 167283	[DNA binding protein (probable) (<i>Staphylococcus subtilis</i>)]	58	25	237
201	1	2669	3307	gi 1511453	[endonuclease III (<i>Methanococcus jannaschii</i>)]	58	34	639
208	1	1	238	gi 1276729	[hycobilisome linker polypeptide (<i>Porphyra purpurea</i>)]	58	29	237
220	11	114575	113058	gi 1397526	[clumping factor (<i>Staphylococcus aureus</i>)]	58	51	1516
231	3	1629	1474	gi 1002520	[Huts (<i>Bacillus subtilis</i>)]	58	45	156
233	6	4201	3497	gi 11463023	[No definition line found (<i>Ctenorhabditis elegans</i>)]	58	19	705
243	10	9203	110082	gi 1537207	[ORF_227 (<i>Escherichia coli</i>)]	58	32	780
257	1	331	1143	gi 1340128	[ORF1 (<i>Staphylococcus aureus</i>)]	58	44	813
302	2	460	801	gi 40174	[ORF X (<i>Bacillus subtilis</i>)]	58	34	342
307	11	6984	6127	gi 1303842	[YQ18 (<i>Bacillus subtilis</i>)]	58	30	858
321	3	1914	2747	gi 1239996	[hypothetical protein (<i>Bacillus subtilis</i>)]	58	41	814
342	4	2724	1497	gi 454638	[ORF 6, putative (<i>Pseudomonas aeruginosa</i>)]	58	41	774
348	1	1	663	gi 1467478	[<i>Unknown</i> (<i>Bacillus subtilis</i>)]	58	36	663
401	2	384	605	gi 143407	[para-aminobenzoic acid synthase, component I (probable) (<i>Bacillus subtilis</i>)]	58	53	222

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
437	1	125	1554	gi 11010866	trps [Bacillus subtilis]	58	35	1220
445	1	105	1442	gi 1581583	protein A [Staphylococcus aureus]	58	32	1338
453	3	789	965	gi 109435	unknown [Schizosaccharomyces pombe]	58	34	177
453	5	2748	2047	gi 1537214	Yjig gene product [Escherichia coli]	58	40	702
479	2	731	1444	gi 1236621	[26.7% identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis]	58	36	714
490	1	909	547	gi 1580920	rood (gtaA) polypeptide (AA 1-473) [Bacillus subtilis] pir S0608 S05048 [probable rood protein - Bacillus subtilis sp P1364 TAGE_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LIPHA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOC)	58	36	363
517	1	1164	85194726 Y018-		HYPOTHETICAL HELICASE NG018.	58	30	1164
517	6	4182	4544	gi 1453422	Orf268 gene product [Hycoplasma hominis]	58	29	363
546	3	2802	4019	gi 1R8052	restriction modification system S subunit [Spiroplasma citri] gi 886052	58	37	1218
562	1	3	179	gi 143831	nfis protein (AA 1-400) [Klebsiella pneumoniae]	58	34	177
600	1	1347	1156	gi 1183039	unknown [Pseudomonas aeruginosa]	58	48	92
604	2	1231	1001	gi 1001253	[hypothetical protein [Synchocystis sp.]	58	41	231
619	1	1	504	gi 1903748	Integral membrane protein [Homo sapiens]	58	43	504
625	1	2	364	gi 1230874	[hypothetical protein [Synchocystis sp.]	58	43	163
635	1	1492	755	gi 11510995	transaldolase [Methanococcus jannaschii]	58	41	738
645	1	1	646	gi 677882	[real sodium-dependent bile acid transporter [Rattus norvegicus] gi 677882	58	33	846
645	3	906	1556	gi 12308999	[hypothetical protein [Bacillus subtilis]	58	41	651
655	1	771	532	gi 1204262	[hypothetical protein (GB:U1028_6) [Haemophilus influenzae]	58	39	240
674	1	615	327	gi 198817	[orfB: homologs to small subunit of phage terminases [Bacillus subtilis]	58	39	309
675	1	1312	806	gi 12181	[orfC gene product [Escherichia coli]	58	28	507
745	1	618	310	gi 1205432	[catalyzes PQ synthesis protein III (pqdIII) [Haemophilus influenzae]	58	32	309
799	2	242	1174	gi 120669	[collagenase [Haemophilus influenzae]	58	36	933
800	2	1096	614	gi 171563	[tRNA isopentenyl] transferase [Saccharomyces cerevisiae] Sp P07881 H03-YEAST tRNA ISOPENTENYLTRANSFERASE (EC 2.5.1.8) ISOPENTENYL-DIPHOSPHATE: tRNA ISOPENTENYLTRANSFERASE (LPP TRANSFERASE) [LPP TRANSFERASE] (LPP)	58	37	483

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
854	1	1108	605	gi 466778	[lysine specific permease [Escherichia coli]	58	44	504
855	1	481	242	gi 861199	[protochlorophyll IX Mg-chalatase subunit precursor [Hordeum vulgare]	58	33	240
891	1	3	527	gi 1293660	[absA2 [Streptomyces coelicolor]	58	31	525
942	1	931	467	gi 405567	[trnH [Plasmid pSK41]	58	30	465
1002	1	952	521	gi 576169	[preLUEK [Staphylococcus aureus]	58	34	432
1438	1	1	261	gi 581558	[isoleucyl tRNA synthetase [Staphylococcus aureus] sp P41368 SY1P_STSAU	58	30	261
					[ISOLEUCYL-TRNA SYNTHETASE MUPROGIN RESISTANT EC 6.1.1.51 (ISOLEUCINE--TRNA LIGASE) (IRES) (MUPROGIN RESISTANCE PROTEIN)]			
1442	1	2	463	gi 971394	[similar to Acc. No. D26185 [Escherichia coli]]	58	34	462
1673	1	480	241	gi 1333951	[amino acid subunit of NADH-dependent glutamate synthase [Electrona boryanum]]	58	38	240
1876	1	3	158	gi 529216	[No definition line found [Candidatus eugensia sp P16501 YXLT_CAEEL]	58	33	156
					[HYPOTHETICAL 7.3 KD PROTEIN P21P12.7 IN CHROMOSOME III.]			
1989	1	108	401	gi 1403458	[YneR [Bacillus subtilis]]	58	29	294
2109	1	3	401	gi 1001801	[hypothetical protein [Synechocystis sp.]]	58	31	399
2473	1	288	145	gi 510140	[11-guandopentapeptide F [Lactococcus lactis]]	58	18	144
2523	1	432	228	gi 644373	[catabolic dehydroquinase dehydratase [Acinetobacter calcoaceticus]]	58	37	225
3041	1	2	211	gi 1203167	[oligopeptide transport ATP-binding protein [Haemophilus influenzae]]	58	39	210
3094	1	3	263	gi 1118288	[isochorismate synthase [Bacillus subtilis]]	58	38	261
3706	1	3	383	gi 155614	[malonate kinase [Archidopsis thalanae]]	58	48	381
3854	1	1	402	gi 808669	[human gcp372 [Homo sapiens]]	58	32	402
4042	1	51	224	gi 508551	[ribulose-1,5-bisphosphate carboxylase large subunit -methyltransferase [Plasmid st11]]	58	37	174
4278	1	3	206	gi 18089	[hypothetical protein [Synchocystis sp.]]	57	31	456
					[hypothetical protein [Escherichia coli]]	57	29	792
31	1	4801	2402	gi 153146	[cerebellar degeneration-associated protein [Homo sapiens]]	57	32	2400
38	14	11611	10796	gi 144659	[orf B [Cleistridium perfringens]]	57	31	816
46	14	12063	13066	gi 1003119	[hypothetical protein [Synchocystis sp.]]	57	25	984

TABLE 2

5 aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (Int)	Stop (Int)	match accession	match gene name	% sim	% ident	length (Int)
51	3	1411	1187	gi 1348516 B318	hypothetical 80K protein - <i>Bacillus sphaericus</i>	57	38	225
54	1	1	1453	gi 164950	listeriolyticus accessory regulator A [<i>Lystophylococcus aureus</i>]	57	31	453
75	1	3	229	gi 11000470	C2787.7 (<i>Caeorhabditis elegans</i>)	57	42	237
92	5	3855	3061	gi 143607	sporulation protein [<i>Bacillus subtilis</i>]	57	35	795
96	3	4006	4773	gi 144297	acetylesterase (Xmc1) [<i>Caldocellum accharolyticum</i>] pir B37202 B37202	57	34	768
107	1	1480	2076	gi 1468955	TtrE [<i>Vibrio cholerae</i>]	57	42	597
109	8	5340	5333	gi 1438846	Unknown [<i>Bacillus subtilis</i>]	57	41	594
112	2	6679	7701	gi 111466250	Unknown [<i>Bacillus subtilis</i>]	57	33	1023
114	4	6384	4108	gi 1871456	putative alpha subunit of formate dehydrogenase [<i>lachnospiraceum</i>]	57	37	2277
126	2	430	1053	gi 28301	lORF2 gene product [<i>Bacillus megaterium</i>]	57	37	624
131	5	6537	6277	gi 1511160	[M. jannaschii] Predicted coding region MJ1163 [<i>Methanococcus jannaschii</i>]	57	36	261
133	3	1268	2201	gi 1303912	[YqW] [<i>Bacillus subtilis</i>]	57	40	468
133	4	1383	2784	gi 1221884	[lureA7] amidolyase [<i>Haemophilus influenzae</i>]	57	37	600
147	4	2164	1694	gi 1467469	Unknown [<i>Bacillus subtilis</i>]	57	33	471
160	2	1293	1060	gi 158604	[lch1] tRNA synthase 2 [<i>Neurospora crassa</i>]	57	26	234
163	8	5687	4764	gi 1145580	[r4D] gene product [<i>Escherichia coli</i>]	57	38	924
168	6	4336	5225	gi 139782	[l33]Da lipoprotein [<i>Bacillus subtilis</i>]	57	32	990
170	5	3397	3455	gi 160304	[Yer161p] [<i>Saccharomyces cerevisiae</i>]	57	37	159
221	6	8026	6809	gi 11316221	carboxypeptidase [Sulfolobus solfataricus]	57	32	1218
228	3	1346	1791	gi 1288569	fibronectin binding protein [<i>Streptococcus dysgalactiae</i>] pir S33850 S33850	57	32	444
263	4	4411	3686	gi 11185002	dihydrodipicolinate reductase [<i>Pseudomonas syringae</i> pv. <i>tabaci</i>]	57	42	726
276	1	494	235	gi 196180	No definition line found [<i>Escherichia coli</i>]	57	40	240
283	2	335	1324	gi 773349	B1B protein [<i>Bacillus subtilis</i>]	57	32	990
297	1	469	236	gi 11334820	reading frame V [<i>Cauliflower mosaic virus</i>]	57	46	234
342	3	1993	2805	gi 1204431	[hypothetical protein] [<i>Haemophilus influenzae</i>]	57	35	813

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (ID)	Stop (ID)	match (nt)	match (nt)	match gene name	match gene name	start (nt)	end (nt)	length (nt)	
375	6	3340	3741	911305177	911305177	[cell division protein [Bacillus subtilis]	[cell division protein [Bacillus subtilis]	57	26	402	
433	6	3286	4011	9111524417	9111524417	[alpha- β -lactamase decarboxylase [Lactococcus lactis]	[alpha- β -lactamase decarboxylase [Lactococcus lactis]	57	40	726	
470	3	503	1145	9111908159	9111908159	[protein serine/threonine kinase [Toxoplasma gondii]	[protein serine/threonine kinase [Toxoplasma gondii]	57	30	243	
487	5	1391	1723	9111507333	9111507333	[ORF1 [Bacillus stearothermophilus]	[ORF1 [Bacillus stearothermophilus]	57	28	333	
498	1	274	852	9111133449	9111133449	[HADH-ubiquinone oxidoreductase subunit 4L [Pedaspora asterinae]	[HADH-ubiquinone oxidoreductase subunit 4L [Pedaspora asterinae]	57	34	579	
503	1	1343	1773	91111502283	91111502283	[organic cation transporter OCT2 [Rattus norvegicus]	[organic cation transporter OCT2 [Rattus norvegicus]	57	30	171	
505	2	1619	1284	9111466884	9111466884	[B1495_c2_194 [Hypobacterium leprae]	[B1495_c2_194 [Hypobacterium leprae]	57	40	336	
519	1	2	1182	2549	91111037207	91111037207	[YrhK [Bacillus subtilis]	[YrhK [Bacillus subtilis]	57	34	1368
522	2	3234	1945	91110644809	91110644809	[homologous to SP-1778A_E001 [Bacillus subtilis]	[homologous to SP-1778A_E001 [Bacillus subtilis]	57	36	1290	
538	2	909	1415	91111517179	91111517179	[phosphoinositide- α -acyltransferase [Streptomyces coelicolor]	[phosphoinositide- α -acyltransferase [Streptomyces coelicolor]	57	40	507	
						[phosphoinositide- α -acyltransferase [Streptomyces coelicolor]	[phosphoinositide- α -acyltransferase [Streptomyces coelicolor]				
547	1	968	1486	9111467340	9111467340	[unknown [Bacillus subtilis]	[unknown [Bacillus subtilis]	57	50	483	
599	1	1062	532	9111026592	9111026592	[PYRANEATE DEHYDROGENASE (EC 1.1.1.12) (PDB: 1NPEH)]	[PYRANEATE DEHYDROGENASE (EC 1.1.1.12) (PDB: 1NPEH)]	57	41	521	
620	2	757	572	9111107894	9111107894	[unknown [Schizosaccharomyces pombe]	[unknown [Schizosaccharomyces pombe]	57	38	146	
622	2	1600	1110	9111172028	9111172028	[thioredoxin 11 [Saccharomyces cerevisiae]	[thioredoxin 11 [Saccharomyces cerevisiae]	57	39	471	
625	2	162	1114	91112623566	91112623566	[hypothetical protein [Hydrobacterium lemnæ]	[hypothetical protein [Hydrobacterium lemnæ]	57	34	753	
680	1	1	204	911143544	911143544	[RNA polymerase sigma-10 factor [Bacillus subtilis] pfr1288251A28635	[RNA polymerase sigma-10 factor [Bacillus subtilis] pfr1288251A28635	57	30	204	
690	1	3	629	911146520	911146520	[transcription initiation factor Sigma H - <i>acillus subtilis</i>	[transcription initiation factor Sigma H - <i>acillus subtilis</i>	57	29	627	
696	1	2	433	9111419972	9111419972	[ipaA8r gene product [Bacillus subtilis]	[ipaA8r gene product [Bacillus subtilis]	57	33	432	
704	1	36	638	9111469931	9111469931	[H. Jannaschii predicted coding region HJ103] [Methanococcus jannaschii]	[H. Jannaschii predicted coding region HJ103] [Methanococcus jannaschii]	57	36	603	
712	1	2316	1621	9111418959	9111418959	[orf4 [Lactobacillus sakei]	[orf4 [Lactobacillus sakei]	57	37	696	
746	1	451	227	9111192973	9111192973	[Rab] [Aplysia californica]	[Rab] [Aplysia californica]	57	42	225	
757	1	20	465	911143979	911143979	[L-curvatus small cryptic plasmid gene for rep protein [Lactobacillus curvatus]	[L-curvatus small cryptic plasmid gene for rep protein [Lactobacillus curvatus]	57	45	447	
862	1	2	295	9111103627	9111103627	[YQ11 [Bacillus subtilis]	[YQ11 [Bacillus subtilis]	57	21	294	
1049	1	907	455	9111510106	9111510106	[ORF-1 [Mycoplasma tumefaciens]	[ORF-1 [Mycoplasma tumefaciens]	57	35	453	
1117	1	1387	695	9111696286	9111696286	[NH2 terminus uncertain [Leishmania tarantolae]	[NH2 terminus uncertain [Leishmania tarantolae]	57	28	693	

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1136	1	2	322	[gi 1130363	[Ycf (Bacillus subtilis)	57	38	321
1144	1	1031	611	[gi 310083	[voltage-activated calcium channel alpha-1 subunit (Rattus norvegicus)	57	46	421
1172	1	1472	736	[gi 1511146	[H] Jannaschii predicted coding region H1113 (Methanococcus jannaschii)	57	28	735
1500	2	746	558	[gi 14270	[putative membrane protein: putative (Bacillus subtilis)	57	15	189
1676	1	659	199	[gi 313777	[uracil permease (Escherichia coli)	57	31	261
2861	1	2	400	[gi 1237015	[orf4 (Bacillus subtilis)	57	23	359
3099	1	3	230	[gi 120540	[isochorismate synthase (Mamophilus influenzae)	57	19	224
3122	1	360	181	[gi 882472	[ORF_0464 (Escherichia coli)	57	40	180
3560	1	2	361	[gi 153490	[tetracycloycin C resistance and export protein (Streptomyces laevigatus)	57	37	360
3850	1	856	434	[gi 155888	[glucose-fructose oxidoreductase (Zymomonas mobilis) [gi 42289] [A42289]	57	40	423
3921	1	704	354	[gi 141353	[glucose-fructose oxidoreductase (EC 1.1.1.1) - precursor - Zymomonas mobilis	57	36	351
3993	1	1	384	[gi 151529	[HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas fluorescens) [pir A44755] [A44756]	57	39	384
4065	1	783	398	[pir JY0037] [RDEC	[nitrate reductase (EC 1.1.1.88) Pseudomonas sp.	57	31	396
4100	1	596	300	[gi 108633	[T06C10.5 gene product (Caenorhabditis elegans)	57	47	297
4163	1	571	2H7	[gi 21512	[patatin (Solanum tuberosum)	57	50	285
4267	2	631	335	[gi 1000165	[spolIIAG (Bacillus subtilis)]	57	38	297
4358	1	3	302	[gi 298032	[EF] (Streptococcus suis)	57	32	300
4389	2	108	290	[gi 40594	[1-phosphofructokinase (Escherichia coli)]	57	37	183
4399	1	2	232	[gi 1484603	[pristinamycin I synthase I (Streptomyces pristinaespiralis)]	57	35	231
4481	1	572	288	[gi 40579	[yeiH (Escherichia coli)]	57	44	285
4486	1	512	25H	[gi 315338	[glutamate synthase (ferredoxin) (Symeocystis sp.) [pir S46957] [S46957]	57	42	255
4510	1	461	242	[gi 1203101	[leukotoxin secretion ATP-binding protein (Hemophilus influenzae)]	57	38	240
4617	1	468	256	[gi 1511222	[restriction modification enzyme, subunit H1 (Methanococcus jannaschii)]	57	35	213
4	11	12201	11524	[gi 149204	[histidine utilization repressor G (Klebsiella aerogenes) [pir A6710] [A36710]	56	31	678
					[formiminoglutamate (EC 3.5.3.8) FORMIMINOGULUTAMATE HYDROLASE (HISTIDINE UTILIZATION PROTEIN G) FRAGMENT]			

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
44	3	1861	2421	gi 4490320	[Y gene product (unidentified)]	56	31	561
44	155	10103	10606	gi 11205099	[hypothetical protein (GB:U19201_11) [Haemophilus influenzae]]	56	39	504
50	6	4820	5161	gi 12039311	[fiber protein (Human adenovirus type 5)]	56	48	342
53	4	2076	2572	gi 622476	[TRANSCRIPTIONAL ACTIVATOR (Providencia stuartii sp P41461 AARP_PROST)]	56	30	697
67	6	5656	6594	gi 466613	[m16A (Escherichia coli)]	56	32	939
89	3	2364	1810	gi 483922	[protein with homology to palI repressor of B. subtilis (Lactobacillus delbrueckii)]	56	39	555
96	1	203	913	gi 1145594	[cAMP receptor protein (tcpP) (Escherichia coli)]	56	35	711
109	21	114250	178446	gi 11204367	[hypothetical protein (GB:U14003_278) [Haemophilus influenzae]]	56	27	405
112	8	5611	6678	gi 1155588	[glucose-fructose oxidoreductase (Zymomonas mobilis) pI A42289 Aa2289]	56	40	1068
131	3	6404	5100	gi 1619724	[fpgE (Bacillus firmus)]	56	30	1305
138	2	65	232	gi 411948	[lpa-2/d gene product (Bacillus subtilis)]	56	31	168
138	4	823	1521	gi 1580866	[lpa-2/r gene product (Bacillus subtilis)]	56	31	699
146	2	740	477	gi 11046009	[M. genitalium predicted coding region M6309 (Mycoplasma genitalium)]	56	37	294
149	2	1639	1067	gi 11943380	[terminase small subunit (Bacteriophage LL-III)]	56	35	573
163	1	2	223	gi 11143947	[glutamine synthetase (Bacteroides fragilis)]	56	30	222
166	5	6745	6449	gi 11403792	[ORF154 (Pseudomonas putida)]	56	26	297
187	1	31	323	gi 1311237	[H(+)-transporting ATP synthase (Zea mays)]	56	30	363
190	1	2	373	gi 11109666	[PROX (Bacillus subtilis)]	56	35	372
191	8	11536	9943	gi 1581070	[acyl coenzyme A synthetase (Escherichia coli)]	56	35	1596
195	1	1291	647	gi 11510242	[collagense (Methanococcus jannaschii)]	56	34	645
230	3	2323	2072	gi 140363	[heat shock protein (Clostridium acetobutylicum)]	56	39	252
238	5	3383	3775	gi 1147533	[ata (Staphylococcus aureus)]	56	31	193
270	2	813	1712	gi 765073	[autolysin (Staphylococcus aureus)]	56	41	900

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

contig	orf	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
120	1	3221	1632	[gi 51751]	[orf1] [Haemophilus influenzae]	56	34	1590
297	5	1140	1373	[gi 151556]	[M. jannaschii] predicted coding region MJ1561 [Methanococcus jannaschii]	56	40	234
321	2	2947	1799	[gi 1001801]	[hypothetical protein] [Synecocystis sp. P-1]	56	31	1149
359	2	1279	641	[gi 46336]	[hypothetical protein] [Haemophilus melioli]	56	26	639
371	2	1360	1823	[gi 145304]	[l-ribulokinase] [Escherichia coli]	56	39	1464
391	4	1762	2409	[gi 1001634]	[hypothetical protein] [Synecocystis sp. P-1]	56	34	648
402	1	380	192	[gi 1138904]	[5-HT1 receptor] [Homo sapiens]	56	48	189
416	4	2480	2109	[gi 11408486]	[HST4A gene product] [Bacillus subtilis]	56	31	372
424	3	1756	2334	[gi 142471]	[acetolactate decarboxylase] [Bacillus subtilis]	56	32	579
457	1	1907	1017	[gi 1205194]	[formamidopyrimidine-DNA glycosylase] [Haemophilus influenzae]	56	36	891
458	2	2423	1812	[gi 115666]	[terminase] [Bacteriophage SP1]	56	37	612
504	2	2152	1283	[gi 112681]	[LppB] [Pasteurella hemolyticus]	56	38	870
511	1	1	1284	[gi 1217049]	[lrrn protein] [Salmonella typhimurium]	56	37	1284
604	3	1099	1701	[gi 1467109]	[rim: J05 ribosomal protein S18 aminine acetyltransferase; 229_C1_170] [Hydrobacterium leprae]	56	43	603
660	5	3547	3774	[gi 1229106]	[2K930_1] [Candidatus elegans]	56	30	228
707	1	35	400	[gi 1151929]	[NADPH-sulfite reductase flavoprotein component] [Salmonella typhimurium]	56	38	266
709	2	1385	1095	[gi 11510801]	[hydrogenase accessory protein] [Methanococcus jannaschii]	56	38	291
718	1	1	495	[gi 141946]	[lpa-24d gene product] [Bacillus subtilis]	56	35	495
744	1	87	677	[gi 1926836]	[repressor protein] [Lactococcus lactis phage B15-7]	56	35	591
790	1	776	399	[gi 11515151]	[ABC transporter, probable ATP-binding subunit] [Methanococcus jannaschii]	56	33	378
795	1	1	3	[gi 11205362]	[cell division protein] [Haemophilus influenzae]	56	34	405
813	1	19	930	[gi 1222161]	[permease] [Haemophilus influenzae]	56	28	912
855	1	3	515	[gi 1226621]	[26.7% of identity in 165 aa to a thermonphilic bacterium hypothetical protein 6; putative] [Bacillus subtilis]	56	33	513
968	1	2	466	[gi 1547513]	[orf3] [Haemophilus influenzae]	56	37	465
973	2	1049	732	[gi 188022]	[Mpr] [Pseudomonas aeruginosa]	56	21	218
1203	1	5	223	[gi 188251]	[HMG-1] [Homo sapiens]	56	34	219

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1976	1	452	237	[gi 19106]	lysine-rich aspartic acid-rich protein [Plasmodium chabaudi] r S22183 S22183 lysine/aspartic acid-rich protein - Plasmodium bradii	56	33	216
2161	1	2	400	[gi 127015]	lORF4 [Bacillus subtilis]	56	27	399
2958	1	362	183	[gi 466685]	No definition line found [Escherichia coli]	56	26	180
2979	1	421	212	[gi 1704354]	spore germination and vegetative growth protein [Haemophilus influenzae]	56	40	210
2994	2	526	126	[gi 866646]	phosphotrioseformylparic ketolosemerase [Rhodobacter phaeocephala]	56	29	201
3026	1	179	328	[gi 143306]	penicillin V amidase [Bacillus sphaericus]	56	30	150
3189	1	289	146	[gi 166604]	Similar to aldehyde dehydrogenase [Caenorhabditis elegans]	56	37	144
3770	1	63	401	[gi 1129145]	Acetyl-CoA Cacyltransferase [Mangifera indica]	56	43	339
4054	1	720	361	[gi 105355]	Na ⁺ /H ⁺ antiporter [Haemophilus influenzae]	56	31	260
4145	1	1	324	[gi 176095]	long-chain acyl-CoA dehydrogenase [Mus musculus]	56	36	124
4200	1	505	254	[gi 1155588]	glucose-fructose oxidoreductase [Zymomonas mobilis] pIC[442289]IA2269 glucose-fructose oxidoreductase (EC 1.1.1.-) reductase - Zymomonas mobilis	56	40	252
4273	1	675	355	[gi 308861]	tGNG start codon [Lactococcus lactis]	56	33	221
1	3	4095	1436	[gi 5341]	Putative orf YCK8c, len:192 [Saccharomyces cerevisiae] [S53559]	55	25	660
11	12	9377	4505	[gi 1216777]	hypothetical protein [Saccharomyces cerevisiae]	55	32	873
12	4	5133	4534	[gi 467337]	halocurate dehalogenase II-1 [Horakella sp.]	55	26	600
19	5	5404	5844	[gi 11001719]	hypothetical protein [Synchocystis sp.]	55	25	441
23	13	14087	112339	[gi 474190]	lucA gene product [Escherichia coli]	55	30	1749
32	7	5368	688	[gi 1340096]	Unknown [Hypothetical protein] [Mycobacterium tuberculosis]	55	37	1521
34	3	2569	1808	[gi 1103968]	YqjQ [Bacillus subtilis]	55	39	762
34	5	3960	3412	[gi 1103962]	YqjK [Bacillus subtilis]	55	33	549
36	1	1291	647	[gi 166045]	lORF_0118 [Escherichia coli]	55	27	645
36	6	6220	5243	[gi 1001341]	Hypothetical protein [Synchocystis sp.]	55	31	978
47	3	3054	3821	[gi 1001819]	Hypothetical protein [Synchocystis sp.]	55	21	768
49	1	2065	1127	[gi 403373]	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pIC[53725]IS1751 glycerophosphoryl diester phosphodiesterase - acillus subtilis	55	36	939

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	Length (nt)	
67	11	8966	9565	[gi 151053]	norA1199 protein (<i>Staphylococcus aureus</i>)	55	23	600	
75	1	861	1273	[gi 16198]	L-histidinol: NAD oxidoreductase (EC 1.1.1.23) (aa 1-334) (<i>Staphylococcus aureus</i>)	55	33	393	
82	9	15387	14194	[gi 1136221]	carboxypeptidase [Sulfolobus solfataricus]	55	35	1194	
87	4	3517	4937	[gi 1064812]	function unknown (<i>Bacillus subtilis</i>)	55	26	1401	
88	2	1172	1636	[gi 182463]	[protein-N(6)-phosphohistidine-sugar phosphotransferase (<i>Escherichia coli</i>)]	55	35	465	
92	1	127	516	[gi 1377832]	function unknown (<i>Bacillus subtilis</i>)	55	36	390	
100	12	836	2035	[gi 1370274]	lzeaxanthin epoxidase (<i>Nicotiana plumbaginifolia</i>)	55	36	1200	
100	5	5137	4558	[gi 396660]	function unknown reading frame (<i>Buchnera aphidicola</i>)	55	29	480	
108	3	4266	2986	[gi 1499866]	[H. Jannasch] predicted coding region MJ1024 (<i>Methanococcus jannaschii</i>)	55	31	1261	
114	3	2616	1834	[gi 1511367]	lformate dehydrogenase, alpha subunit (<i>Methanococcus jannaschii</i>)	55	29	783	
144	3	1805	1476	[gi 1100767]	function unknown (<i>Saccharomyces cerevisiae</i>)	55	35	330	
165	5	6212	5508	[gi 105884]	[H. genitalium] predicted coding region HG199 (<i>Hyphomicrobium genitalium</i>)	55	27	705	
189	5	2205	2376	[gi 142569]	[ATP synthase α subunit (<i>Bacillus firmus</i>)]	55	35	372	
191	6	9136	6857	[gi 59411]	[B0372_3 (<i>Caenorhabditis elegans</i>)]	55	39	280	
194	2	264	616	[gi 1165768]	K7 kinesin-like protein (<i>Dictyostelium discoideum</i>)	55	34	273	
209	4	1335	1676	[gi 147357]	[chi4 gene product (<i>Schizosaccharomyces pombe</i>)]	55	35	342	
211	2	1693	1145	[gi 141030]	ORF6 (<i>Bacillus subtilis</i>)	55	37	549	
213	1	2	644	[gi 1372]	[gi 1633692]	TRA (Yersinia enterocolitica)	55	28	729
214	7	4144	5481	[gi 1001793]	hypothetical protein (<i>Synechocystis sp.</i>)	55	10	1338	
221	7	11473	9197	[gi 466520]	poCR (<i>Salmonella typhimurium</i>)	55	32	2277	
233	6	5906	4817	[gi 1237063]	Unknown (<i>Mycobacterium tuberculosis</i>)	55	18	1092	
236	4	1375	2340	[gi 11616199]	putative (<i>Bacillus subtilis</i>)	55	32	966	
243	2	380	1885	[gi 455907]	mercuric reductase (Plasmid p1258)	55	29	1506	
258	1	786	394	[gi 455006]	[gi 6 (Rhodococcus fascians)]	55	36	193	
281	1	126	938	[gi 140893]	homologous to SwissProt YIDA_ECOLI hypothetical protein (<i>Bacillus subtilis</i>)	55	35	813	
316	1	1323	2102	[gi 188647]	[luxA homologue (<i>Rhizobium</i> sp.)]	55	30	780	
326	5	2968	2744	[gi 1396324]	[proline iminopeptidase (<i>lactobacillus halveticus</i>)]	55	16	225	

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start ID (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
351	2	2322	1429	gi 11206820	hydrogen peroxide-inducible activator [Haemophilus influenzae]	55	24	894
353	4	2197	2412	gi 12272475	chitin synthase (Emericella nidulans)	55	50	216
380	1	14	379	gi 142554	ATP synthase I subunit [Bacillus megaterium]	55	37	266
383	1	462	232	gi 1827272	tertiochrome-binding protein [Bacillus subtilis]	55	36	231
386	1	3	938	gi 11510251	DNA helicase, putative [Haemophilus jannaschii]	55	30	936
410	1	1208	1891	gi 11205144	multidrug resistance protein [Haemophilus influenzae]	55	27	684
483	2	411	833	gi 113934	lipo-10r gene product [Bacillus subtilis]	55	26	423
529	3	1777	1433	gi 606150	ORF f109 [Escherichia coli]	55	33	345
555	1	1088	585	gi 143407	[para-aminobenzoic acid synthase, component 1 (parB) [Bacillus subtilis]]	55	28	504
565	1	402	202	gi 1223961	ICD-tyrosine epimerase [Versinia pseudotuberculosis]	55	41	201
582	1	751	452	gi 1256663	20.2% identity with NADH dehydrogenase of the <i>Leishmania major</i> mitochondrial; putative [Bacillus subtilis]	55	36	300
645	5	2260	2057	gi 110824	[fusion protein F (bovine respiratory syncytial virus) pif J01481 VGRBAA fusion glycoprotein precursor - bovine respiratory syncytial virus (strain A31908)]	55	25	204
672	2	957	2216	gi 11511333	N. jannaschii predicted coding region NJ1322 [Methanococcus jannaschii]	55	36	1260
710	1	955	479	gi 1537007	ORF_1379 [Escherichia coli]	55	30	477
737	1	1859	945	gi 1536763	[CG site No. 18166] [Escherichia coli]	55	30	915
742	2	228	572	gi 130160	[product unknown [Bacillus subtilis]]	55	18	345
817	2	1211	903	gi 11136289	histidine kinase A [Dicytostelium discoideum]	55	29	309
819	1	582	352	gi 155073	[polymorphic antigen [Plasmodium falciparum]	55	22	220
832	2	1152	724	gi 40367	ORFC [Clostridium acetobutylicum]	55	32	429
840	1	769	386	gi 1205895	pseudouridylate synthase 1 [Haemophilus influenzae]	55	39	384
1021	1	23	529	gi 48563	beta-lactamase [Verminia enteroclitica]	55	38	507
1026	1	60	335	gi 47604	Opp C (M1-31) [Salmonella typhimurium]	55	26	276
1525	1	1	282	gi 1477531	[ata] [Staphylococcus aureus]	55	29	282
1814	2	224	985	gi 1046078	1H. genitalium predicted coding region HG3169 [Mycoplasma genitalium]	55	38	762
3254	1	427	254	gi 413968	lipo-44d gene product [Bacillus subtilis]	55	30	174

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

5
10
15
20
25
30
35
40
45
50

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	sim	ident	length (nt)
3695	1	686	345	gi 216771	haloacetate dehalogenase H-1 [Moraxella sp.]	55	32	342
3721	1	1	312	gi 142029	ORF1 gene product [Escherichia coli]	55	31	312
3759	1	3	272	gi 142029	ORF1 gene product [Escherichia coli]	55	36	270
3869	1	22	423	gi 1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	55	45	402
3916	1	2	385	gi 1529754	lspc [Streptococcus pyogenes]	55	38	384
3955	1	4	198	gi 1476252	lphase 1 flagellin [Salmonella enterica]	55	36	195
4074	1	488	216	gi 42029	ORF1 gene product [Escherichia coli]	55	38	243
4164	1	2	343	gi 1524267	unknown [Mycobacterium tuberculosis]	55	28	342
4284	1	14	208	gi 1100774	ferredoxin-dependent glutamate synthase [Synechocystis sp.]	55	36	195
4457	2	644	378	gi 180189	cerebellar degeneration-related antigen [CDR34] [Homo sapiens] gi 162737	55	38	267
					cerebellar degeneration-associated protein [Homo sapiens] pir A29770 A29770 cerebellar degeneration-related protein - human			
4514	1	2	344	gi 1216773	haloacetate dehalogenase II-1 [Moraxella sp.]	55	32	243
4559	1	432	217	gi 1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	55	42	216
4606	1	416	210	gi 186120	myosin alpha heavy chain (S2 subfragment) [rabbit], myosin, actin	55	27	207
5	n	5348	4932	gi 1536059	ORF YBL07c [Saccharomyces cerevisiae]	54	27	417
12	7	7166	6165	gi 120504	homoserine acetyltransferase [Mamophilus influenzae]	54	30	1002
23	16	17086	15326	gi 470192	lucC gene product [Escherichia coli]	54	31	1761
35	1	2	979	gi 148054	small subunit of soluble hydrogenase (aa 1-38) [Synechococcus sp.] gi s06919 HQ0455 soluble hydrogenase (EC 1.12.1.-) small chain -	54	36	978
					methococcus sp. (PCG 6716)			
37	11	9437	8667	gi 537107	ORF_E277 [Escherichia coli]	54	38	771
37	12	8165	8332	gi 1160967	1beta/1alpha-1-protein thioesterase [Homo sapiens]	54	37	168
46	15	13025	13004	gi 438173	protein is hydrophobic, with homology to E. coli PROW: putative Bacillus subtilis	54	28	780
56	2	263	736	gi 1126139	fbpB [Bacillus subtilis]	54	34	534
57	13	11117	11079	gi 1151248	inosine-uridine preferring nucleoside hydrolase [Citrithidium fasciculata]	54	32	939
66	2	516	1133	gi 1333781	Cap [Drosophila melanogaster]	54	29	618
70	10	8116	8446	gi 1395823	lphoE [Rhizobium meliloti]	54	31	531

TABLE 2

5
10
15
20
25
30
35
40
45
50
55

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (Int)	Stop (Int)	match accession	match gene name	sim	ident	length (Int)
70	112556	116801	116801	sp P02883 TCR_S	METACYCLINE RESISTANCE PROTEIN	54	29	736
87	5	4915	5706	gi 1106811	[function unknown (Bacillus subtilis)]	54	33	792
92	4	3005	2289	gi 1203366	[oligopeptide transport ATP-binding protein (Haemophilus influenzae)]	54	33	717
103	2	2596	1556	gi 171095	[protein kinase (Bacillus brevis)]	54	33	1041
105	2	3585	2095	gi 143727	[putative (Bacillus subtilis)]	54	30	1391
112	1	2117	2272	gi 153724	[halic (Streptococcus pneumoniae)]	54	41	396
127	2	1720	2493	gi 144237	[acetyl esterase (XynC) (Caldobacterium saccharolyticum) pir B37203 B37202	54	34	774
138	1	1600	3106	gi 142711	[pyruvate oxidase (Escherichia coli)]	54	36	1207
152	2	525	1172	gi 1377834	[unKnown (Bacillus subtilis)]	54	23	648
161	1	4831	5669	gi 190305	[ORF73 (Bacillus subtilis)]	54	28	639
161	13	6694	7251	gi 1511039	[phosphate transport system regulatory protein (Methanococcus jannaschii)]	54	32	558
164	6	3263	4543	gi 1204976	[polyU-tRNA synthetase (Haemophilus influenzae)]	54	34	1281
164	120	21602	22243	gi 1143582	[spolIIE protein (Bacillus subtilis)]	54	32	642
171	6	5683	4250	gi 1435965	[fmlA] gene products (Bacillus stearothermophilus) pir S43314 S43914	54	37	1434
206	118	13208	13720	gi 11240016	[R0E013 (Caenorhabditis elegans)]	54	36	513
218	1	1090	1905	gi 1467378	[unKnown (Bacillus subtilis)]	54	26	816
220	1	1322	663	gi 1353761	[myosin II heavy chain (Naegleria fowleri)]	54	22	660
220	113	112655	113059	pir S0085 S004	[gene 11-1 protein precursor - Plasmidone fMetparon (fragment)]	54	35	405
221	1	2030	3709	gi 11130811	[YqM (Bacillus subtilis)]	54	34	1680
272	7	5055	4219	gi 162964	[arylamine N-acetyltransferase (AA 1-290) (Gallus gallus) pir S06652 YCHY3	54	33	837
316	1	4141	4701	gi 162769	[mcE gene product (Escherichia coli)]	54	31	561
316	110	6994	8742	gi 413951	[Ipa-2/3 gene product (Bacillus subtilis)]	54	28	1749
338	1	3377	2214	gi 490328	[ORF F (unidentified)]	54	28	1164
341	1	3201	3614	gi 171959	[myosin-like protein (Saccharomyces cerevisiae)]	54	25	414

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (Int)	Stop (Int)	match accession	match gene name	% sim	% ident	length (nt)
346	1	1620	912	[gi 396400	similar to eukaryotic Na+/H+ exchangers (Escherichia coli) SP P32703 YCE_ECOLI_HYPOTHETICAL_60.5_KD_PROTEIN_IN_SOMK-ACS_NTEROGENIC_REGION_(054).	54	34	909
346	2	623	1351	[gi 537109	[ORF_F346 (Escherichia coli)]	54	34	729
378	2	1007	1942	[sp P02981 PCR_S	TETRACYCLINE RESISTANCE PROTEIN.	54	31	936
408	6	4351	5201	[gi 474190	lucA gene product (Escherichia coli)	54	29	951
444	9	7934	8854	[gi 216267	[ORF2 (Bacillus megaterium)]	54	32	921
463	2	2717	2229	[gi 101660	product unknown (Bacillus subtilis)	54	50	489
502	2	1696	1133	[gi 1105015	hypothetical protein (SP P10120) [Haemophilus influenzae]	54	38	564
503	6	6262	5357	[gi 1505558	[2-hydroxyhepta-2,4-diene-1,7-diole isomerase (Methanococcus jannaschii)]	54	41	906
550	1	2736	1522	[gi 40100	rodC (tagB) polypeptide (AA 1-746) [Bacillus subtilis] [R] [sp S0649 sp0649	54	35	1215
					rodC protein - Bacillus subtilis [P11485 TAGP_BACSU_TECHNIC_ACID_BIOSYNTHESIS_PROTEIN_F.			
551	5	JJ05	4279	[gi 350197	lunknow (Corynebacterium glutamicum)	54	34	975
558	2	1156	958	[gi 485090	No definition line found [Caenorhabditis elegans]	54	32	399
580	1	91	916	[gi 311906	[fused envelope glycoprotein precursor (Friend spleen focus-forming virus)]	54	45	846
603	1	554	757	[gi 1233423	[ORF_VGR34W (Saccharomyces cerevisiae)]	54	36	204
617	1	25	249	[gi 219959	[ornithine transcarbamoylase (Homo sapiens)]	54	40	225
622	3	1097	1480	[gi 11303873	[YopZ (Bacillus subtilis)]	54	25	384
623	1	3	404	[gi 1083250	low homology to P20 protein of <i>Bacillus licheniformis</i> and bleomycin acetyltransferase of <i>Streptomyces verticillus</i> (Bacillus subtilis)	54	45	402
689	1	1547	1011	[gi 52446	[NADH dehydrogenase subunit 4 (Apis mellifera ligustica) sp S2968 S523968	54	30	537
725	2	686	1841	[gi 987096	[NADH dehydrogenase chain 4 - honeybee mitochondrion (SGC4)]	54	26	756
956	1	1	269	[pir S0702 S307	sensory protein kinase [Streptomyces hygroscopicus]	54	24	249
					integrin homolog - Yeast (Saccharomyces cerevisiae)			
978	2	1137	859	[gi 101994	[ORF_YNL012w (Saccharomyces cerevisiae)]	54	33	279
1114	1	3	281	[gi 1001108	hypothetical protein [Synchocystis sp.]	54	33	279
2450	1	1	228	[gi 050537	Ich-TOG (Homo sapiens)	54	32	228
2934	1	1	387	[gi 508870	[sp 37d_008 gene product: (Bacillus subtilis)]	54	36	187
2970	1	499	251	[sp P774P YCE_	HYPOTHETICAL PROTEIN IN ASPS 5 REGION (FRAGMENT).	54	42	249

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3002	1	1	309	[gi 44027]	[Tau protein (Lactococcus lactis)]	54	33	309
3561	1	9	664	[gi 151259]	[HMG-CoA reductase (EC 1.1.1.86) (Pseudomonas mavalonitii pir A44756 A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.86) Pseudomonas sp.]	54	35	456
3572	1	72	401	[gi 450688]	[hdc gene of Ecopri gene product [Escherichia coli pir S1837 S09629 hypothetical protein A - Escherichia coli (SUB 40-520)]	54	16	330
3829	1	798	400	[gi 1122245]	[malavonate pyrophosphate decarboxylase (Rattus norvegicus)]	54	29	399
3909	1	1	273	[gi 29865]	[CNP-E (Homo sapiens)]	54	30	273
3921	1	3	209	[pir S24325 S24]	[glucan 1,4-beta-glucosidase (EC 3.2.1.74) - Pseudomonas fluorescens subsp. cellulosa]	54	34	207
4438	1	566	285	[gi 1119657]	[unknown protein (Mycoplasma pneumoniae)]	54	30	282
4459	1	3	272	[gi 1046081]	[hypothetical protein (GB:D06185_10) (Mycoplasma genitalium)]	54	38	270
4564	1	3	221	[gi 216267]	[ORF2 [Bacillus megaterium]]	54	38	219
23	112	112538	10685	[gi 474192]	[lucC gene product (Escherichia coli)]	53	35	1854
23	14	114841	113579	[gi 42029]	[orf1 gene product (Escherichia coli)]	53	32	1263
24	3	4440	3940	[gi 11169947]	[c2 gene product (Bacteriophage 61)]	53	36	501
26	4	3818	4618	[gi 11486247]	[unknown [Bacillus subtilis]]	53	37	801
38	6	12856	1998	[gi 405810]	[orf1 (Escherichia coli)]	53	40	1143
38	10	9380	7806	[gi 1139954]	[thyroid sodium/iodide symporter NIS (Rattus norvegicus)]	53	29	1575
56	10	112324	112100	[pir A5492 A545]	[110k actin filament-associated protein - chicken]	53	32	225
57	6	5047	4583	[pir A00141 DEP]	[alcohol dehydrogenase (EC 1.1.1.1) - fission yeast (Schizosaccharomyces pombe)]	53	39	465
57	12	110515	8932	[gi 1480129]	[putative transcriptional regulator [Bacillus stearothermophilus]]	53	30	1584
67	12	9496	10218	[gi 1511555]	[quinolone resistance nodA protein [Methanococcus jannaschii]]	53	31	721
69	3	3125	2302	[gi 11687017]	[arabinogalactan-protein, AGP [Nicotiana alata, cell-suspension culture filtrate, peptide, 461 a.a.]]	53	30	744
79	1	3	1031	[gi 1523602]	[Glucanase (Anabaena variabilis)]	53	32	1029
80	1	673	338	[gi 452428]	[NTPase 3 [Plasmodium falciparum]]	53	36	336
88	4	1910	2524	[gi 17034]	[ORF_0488 [Escherichia coli]]	53	25	615
88	5	2467	3282	[gi 1517034]	[ORF_0488 [Escherichia coli]]	53	29	816

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
52	8	5870	5505	[gi 393958]	lymphotropic murine retrovirus receptor [<i>Rattus norvegicus</i>]	53	33	366
94	5	4417	3229	[gi 173038]	trypomosin (TPM1) [<i>Saccharomyces cerevisiae</i>]	53	25	1179
99	5	4207	5433	[sp PA8246 BCR_E]	BICYCLICIN RESISTANCE PROTEIN (SULFONAMIDE RESISTANCE PROTEIN).	53	30	1227
120	3	1639	2262	[gi 176655]	[ORF1] [<i>Vibrio anguillarum</i>]	53	35	624
120	11	7257	8897	[gi 1524397]	glycine betaine transporter Opd1 [<i>Bacillus subtilis</i>]	53	33	1641
127	6	6893	5685	[gi 1256630]	[putative] <i>Bacillus subtilis</i>	53	32	1209
147	2	255	557	[gi 1881648]	[epis gene product] [<i>Staphylococcus epidermidis</i>]	53	34	101
158	4	4705	4256	[gi 151004]	Imidoyl regulatory protein in AlgR [<i>Pseudomonas aeruginosa</i>] pir [32B02 IA2802]	53	32	450
					regulatory protein AlgR - <i>Pseudomonas aeruginosa</i> sp P26272 ALGR_PSEAE			
					POSITIVE ALGINATE BIOSYNTHESIS REGULATORY PROTEIN.			
171	7	5717	5421	[gi 1510669]	hypothetical protein (sp D64044_18) [<i>methanococcus jannaschii</i>]	53	34	297
191	9	113087	11483	[gi 2980885]	acetoacetate decarboxylase [Clostridium acetobutylicum] pir [89346 BA9346]	53	31	1605
					butyrate-acetoacetate CoA-transesterase (EC 8.3.9) small chain -			
					Clostridium acetobutylicum sp P13752 CFTA_CLOAB BUTYRATE-ACETOACETATE COA- TRANSFERASE SUBUNIT (EC 2.8.3.9) (COAT A)			
203	5	3763	4326	[gi 143156]	[tpo protein (tpo start codon)] [<i>Bacillus subtilis</i>]	53	29	564
206	17	18204	18971	[gi 304136]	acetylglutamate kinase [aerobacter/thermophilus] sp P007905 ARGB_BACST	53	36	768
					ACETYLGLUTAMATE KINASE (EC 2.7.2.8) (NAG kinase) (NAG kinase) IN-ACETYL-L- GLUTAMATE 5-PHOTOTRANSFERASE.			
212	10	4021	4221	[gi 9878]	protein kinase [Plasmodium falciparum]	53	24	261
231	2	1580	1350	[gi 157506]	[paromycin] [<i>Dirofilaria immitis</i>]	53	34	231
272	6	2719	3249	[pir A33141 A331]	hypothetical protein (tpoD 3' region) - <i>Streptococcus mutans</i>	53	34	531
308	3	927	2576	[gi 106292]	[ORF_0696] (<i>Escherichia coli</i>)	53	33	1650
320	7	5645	5884	[gi 160596]	RNA polymerase III large subunit [Plasmidium falciparum]	53	33	240
					sp P21625 RPC1_PLAFA DNA-DIRECTED RNA POLYMERASE III LARGEST SUBUNIT (EC 2.7.7.6).			
327	1	218	901	[gi 834601]	[unknown] [<i>Schizosaccharomyces pombe</i>]	53	31	684
341	2	212	2500	[gi 1613732]	[ORF1] [<i>Campylobacter jejuni</i>]	53	31	2289
351	1	763	1383	[sp P31675 YABH]	[HYPOTHETICAL 42.7 KD PROTEIN IN TPA-LEUD INTERGENIC REGION (ORF04)]	53	32	381
433	7	5087	4731	[gi 1001961]	[HNC class II analog] [<i>Staphylococcus aureus</i>]	53	30	357
454	2	1240	980	[pir A60328 A303]	40k cell wall protein precursor (ter S' region) - <i>Streptococcus mutans</i>	53	27	261
					strain OM175, serotype f)			

TABLE 2

EP 0 786 519 A2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start ID	Stop ID	match Int	Accession	orfQ gene name	% sim	% ident	length (nt)
470	4	1123	1761	1911518246	rat CCP360 [Rattus rattus]		53	30	639
483	1	432	217	911440429	putative transcriptional regulator [Bacillus stearothermophilus]		53	23	216
544	1	516	1259	911446887	ORF 1 (AA 1 - 121) (1 is 2nd base in codon) [Staphylococcus aureus] ORF 1 (615765 S15765 hypothetical protein 1 (rib 5' region) - <i>staphylococcus</i> <i>aureus</i> (fragment)		53	38	744
558	10	3957	3754	911151440	res gene [Bacteriophage PI]		53	32	204
603	2	339	620	19150738	[<i>Vibrio parahaemolyticus</i>] toxin shock syndrome toxin-1 precursor [Staphylococcus aureus] phi[24606 XGSAS1] toxic shock syndrome toxin-1 precursor - <i>staphylococcus</i> <i>aureus</i>		53	26	282
693	1	1669	941	911151323			53	38	729
766	1	2	673	1911687600	orfA2: orfA2 forms an operon with orfA1 (<i>Listeria monocytogenes</i>)		53	43	672
781	1	667	315	9111204551	[<i>Hemophilus influenzae</i>] plin biogenesis protein [Hemophilus influenzae]		53	26	333
801	1	3	545	9111219400	[<i>Escherichia coli</i>] SAPB protein [Escherichia coli]		53	25	543
803	1	2	910	911693278	lipase-like enzyme [Alcaligenes eutrophus]		53	30	909
872	1	1177	550	911290312	[<i>Streptococcus suis</i>] EF (Streptococcus suis)		53	30	588
910	1	2	164	9111049316	unknown [Schizosaccharomyces pombe]		53	29	183
943	1	794	399	911190508	similar to unidentified ORF near 47 minutes (<i>Escherichia coli</i>) sp P1116 YICK_ECOLI_HYPOTHETICAL 43.5 KD PROTEIN IN SELC-NLPA_NTERGENIC REGION.		53	30	396
988	1	1004	504	911142441	[<i>Bacillus subtilis</i>] ORF 3: putative (<i>Bacillus subtilis</i>)		53	28	501
1064	1	3	434	91130080	[<i>Entamoeba histolytica</i>] myosin heavy chain		53	26	432
1366	1	3	452	911308852	[<i>Lactococcus lactis</i>] transmembrane protein [Lactococcus lactis]		53	33	450
1758	1	792	397	9111001774	hypothetical protein [Synchocystis sp. P-1]		53	30	396
1837	1	1	447	9111303919	[<i>Yersinia</i>] YqjX [Bacillus subtilis]		53	27	447
2381	1	798	400	911116243	22.4% identity with <i>Escherichia coli</i> DNA-damage inducible protein ... putative (<i>Bacillus subtilis</i>)		53	37	399
3537	1	1	327	911450688	hsdK gene of Ecoprr1 gene product (<i>Escherichia coli</i>) pi S8B17 S138437 hsdK protein - <i>Escherichia coli</i> pi S09629 S09629 hypothetical protein A - <i>Escherichia coli</i> (SUB 40-520)		53	35	327
3747	2	137	397	9111477486	[<i>Burkholderia cepacia</i>] transposase (<i>Burkholderia cepacia</i>)		53	53	261
11	5	3049	3441	1911863224	No definition line found (<i>Ceenorhabditis elegans</i>)		52	33	393

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	Length (nt)
15	5	2305	2369	gi 215966	G41 protein (gtg start codon) [Bacteriophage T4]	52	34	165
19	3	2429	3608	gi 1205319	UDP-murac-pentapeptide synthetase [Haemophilus influenzae]	52	31	1380
24	1	6920	3462	gi 1579124	predicted 86-kDa protein; 52% observed [Mycobacteriophage 15] pir S30871 S10971 gene 26 protein - Mycobacterium phage 15 sp 005223 VG26_BPM15_MINOR TAIL PROTEIN_QP26, (SUB 2-617)	52	32	3459
37	5	3015	3935	gi 1500543	P115 protein [Methanococcus jannaschii]	52	25	921
38	13	8795	9703	gi 46831	glucose kinase [Streptomyces coelicolor]	52	29	909
44	16	110617	111066	gi 42012	InrAE gene product [Escherichia coli]	52	36	450
46	1	3	521	gi 1 0 0957	[NADH dehydrogenase subunit 6 [Anopheles trinkaei]	52	25	519
51	10	5531	6280	gi 182269	trc [Plasmid pAD1]	52	32	750
56	5	3968	2826	gi 181949	endothelial differentiation protein (edg-1) [Homo sapiens] pir A35300 A15300 C protein-coupled receptor edg-1 - human sp P21451 EDG1_HUMAN PROBABLE G PROTEIN-COUPLED RECEPTOR EDG-1.	52	23	1143
57	5	4850	4173	gi 104153	sorbitol dehydrogenase [Bacillus subtilis]	52	27	678
62	5	3364	2870	gi 1 0 2199	[phoE gene product [Bacillus subtilis]	52	25	495
62	6	4445	3651	gi 1 6 485	[NADH dehydrogenase [Synechococcus PCC7942]	52	27	795
67	14	11155	12962	gi 12511365	[glutamate synthase (NADPH), subunit alpha [Methanococcus jannaschii]	52	30	1608
67	21	116915	18158	gi 1204391	[hypothetical protein (SP_P1122) [Haemophilus influenzae]	52	25	1224
70	4	2185	1937	gi 17227	cytoplasmic dynein heavy chain [Dictyostelium discoideum] r A44357 AA44357	52	16	189
					dynein heavy chain, cytosolic - slime mold cytoskeleton discoleidium			
96	10	110005	10664	gi 111408465	B65G gene product [Bacillus subtilis]	52	26	660
103	5	3986	3351	gi 1 0 9368	[respiratory nitrate reductase (Bacillus subtilis)]	52	42	636
109	3	4102	3350	gi 1693274	lmbE gene product [Mycobacterium leprae]	52	39	753
109	19	15732	17700	gi 1526981	lmino acid permease YaeF like protein [Salmonella typhimurium]	52	30	1569
121	3	1412	1981	gi 1 7 2831	[Unknown [Saccharomyces cerevisiae]]	52	32	432
125	3	865	1660	gi 1296975	[put gene product [Porphyromonas gingivalis]	52	38	816
130	2	659	1807	gi 1256634	[25.8% identity over 120 aa with the Synecoccus sp. spv protein: putative [Bacillus subtilis]]	52	36	1149
149	1	1164	583	gi 1225913	[PBX terminase [Bacillus subtilis]]	52	33	582
149	14	4687	4415	gi 1510368	[H. jannaschii predicted coding region KJ0272 [Methanococcus jannaschii]]	52	35	273

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (10 Int)	Stop (Int)	match accession	match gene name	sim	ident	length (nt)
167	1	216	1001	gi 1146025	cell division protein [Escherichia coli]	52	43	786
188	1	120	1256	gi 1474915	orf 337; translated orf similarity to SW: BCR_ECOLI bicistronic resistance protein of Escherichia coli (Coxiella burnetii) pir SA4207 SA4207 hypothetical protein 337 - Coxiella burnetii (sub -338)	52	26	1137
195	9	9161	8760	gi 1028	mitochondrial outer membrane 72K protein [Neurospora crassa] P A3682 A3682 72K mitochondrial outer membrane protein - <i>Neurospora crassa</i>	52	25	402
200	1	2065	2607	gi 142439	ATP-dependent nuclease (Bacillus subtilis)	52	35	543
203	4	2776	3684	gi 1303698	BLT2 [Bacillus subtilis]	52	25	909
227	8	5250	5651	gi 105080	lysoin heavy chain [Entamoeba histolytica]	52	24	402
242	1	21	1424	gi 1050877	[Bacillus coli]	52	32	104
249	5	4526	4753	gi C7222 C372	cytochrome P450 1A1 hepatic - dog (fragment)	52	23	228
255	1	2107	1055	gi 143290	penicillin-binding protein [Bacillus subtilis]	52	28	1053
276	7	3963	3664	gi 1001610	hypothetical protein [Synechocystis sp.]	52	10	100
276	8	4456	4055	gi 146235	orf 13 [Hycoplasma capricolum]	52	26	402
289	2	1856	1449	gi 150900	ATP phosphodiesterase [Proteus vulgaris]	52	34	408
325	1	1	279	gi 1204874	polypeptide deformylase (formylmethionine deformylase) [Haemophilus influenzae]	52	33	279
340	1	2017	1010	gi 1215695	peptide transport system protein SapF homolog; SapF homolog [Hycoplasma pneumoniae]	52	33	1008
3757	3	340	1878	gi 1467446	similar to SpvB [Bacillus subtilis]	52	28	1539
424	4	4104	3262	gi 1478239	[Unknown [Hycobacterium tuberculosis]]	52	34	843
430	1	3	575	gi PA2606 A426	orf 5' to orf 405 - Sacccharopolypora erythraea (fragment)	52	28	573
444	4	4728	3712	gi 1408494	homologous to penicillin acylase [Bacillus subtilis]	52	31	1017
465	1	1802	903	gi 143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir A37630 A27630 regulatory protein phoR - Bacillus subtilis SP P23545 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN NOR (EC 2.7.3.-)	52	36	900
469	5	4705	4169	gi 1755152	highly hydrophobic Integral membrane protein [Bacillus subtilis] SP PA2953 TAGC_BACSU TRICHOIC ACID TRANSLLOCATION PERMEASE PROTEIN AGC,	52	32	537
495	1	1262	633	gi 1204607	transcription activator [Hemophilus influenzae]	52	25	610
505	7	6004	5762	gi 142440	ATP-dependent nuclease [Bacillus subtilis]	52	26	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
517	2	1162	1614	gi 166162	Bacteriophage phi-11 int gene activator [Staphylococcus acetylphage phi 11]	52	35	453
543	2	444	1295	gi 1215693	[putative orf; Gt9_orf14] [Mycoplasma pneumoniae]	52	25	852
586	1	1	336	gi 1581648	[epibio gene product [Staphylococcus epidermidis]	52	36	336
771	1	848	426	gi 1279769	[FdHC [Methanohalobacterium thermoformicium]	52	30	423
1120	12	100	330	gi 142439	[ATP-dependent nuclelease [Bacillus subtilis]	52	35	231
1614	1	691	347	gi 1282626	[CME ORF3] [Bacillus subtilis]	52	26	345
2495	1	1	324	gi 1216151	DNA polymerase (gene L; int start codon) [Bacteriophage SP02] gi 579197 SP02 DNA polymerase (aa 1-648) [Bacteriophage SP02] pir A21698 DJBPS2 DNA-directed DNA polymerase (EC 2.7.7.1) - phage P02	52	34	324
2931	1	566	285	gi 1226636	[YbbG [Bacillus subtilis]	52	30	282
2943	1	577	120	gi 111713	[hla ORF (AA 1-265) [Escherichia coli]	52	35	258
2953	1	588	295	gi 1298032	[EF [Streptococcus suis]	52	34	294
3667	1	612	307	gi 1849025	[hypothetical 64.7-kDa protein [Bacillus subtilis]	52	36	306
3944	1	478	260	gi 1218040	[BAA [Bacillus licheniformis]	52	36	219
3954	2	613	347	gi 185066	[U87 [Human herpesvirus 6]	52	50	267
3966	1	90	401	gi 1205919	[Na ⁺ and Cl ⁻ -dependent gamma-aminobutyric acid transporter [Bacillus subtilis]	52	33	312
4002	1	3	189	gi 140003	[influenzae hydrogenase (NAD ⁺) I [Bacillus subtilis] gi P231129 DOD1-BACU2 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2.2) [ALPHA-KETOGLUTARATE DEHYDROGENASE].	52	42	387
4020	1	1	249	gi 1152188	[ornithine decarboxylase [Leishmania donovani]]	52	47	249
4038	1	438	220	gi 1409795	[No definition line found [Escherichia coli]]	52	32	219
4248	1	3	212	gi 1963077	[Adr6p [Saccharomyces cerevisiae]	52	40	210
7	1	3	575	gi 893747	[putative cel operon regulator [Bacillus subtilis]	51	28	573
21	4	2479	3276	gi 1510962	[indole-3-glycerol phosphate synthase [Methanococcus jannaschii]]	51	32	798
22	9	5301	5966	gi 103933	[YQIN [Bacillus subtilis]]	51	25	666
43	1	1516	1283	gi 1519460	[Srp1 [Sclerotium accharomyces pombe]]	51	31	234
44	17	11042	111305	gi 142011	[mabD gene product [Escherichia coli]]	51	35	264
51	11	6453	6731	gi 1494471	[vacuolating toxin [Helicobacter pylori]]	51	37	279

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (Int)	Stop (Int)	match accession	match gene name	% sim	% ident	length (nt)
52	4	2537	2995	gi 1236632	258 [identity to the E.coli regulatory protein MprA; putative [Bacillus subtilis]	51	32	459
57	10	7331	6843	gi 508173	EL1A domain of PTS-dependent Gat transport and phosphorylation Escherichia coli	51	32	489
59	1	29	1111	gi 1293163	alanine dehydrogenase [Bacillus subtilis]	51	33	1083
67	120	115791	105876	gi 11510977	M. jannaschii predicted coding region K40938 [Methanococcus jannaschii]	51	24	786
69	2	11559	11218	gi 1467359	unknown [Bacillus subtilis]	51	34	342
71	1	3	1196	gi 1298032	EF [Streptococcus suis]	51	32	1194
78	2	1349	176	gi 1111242	proliferating cell nuclear antigen [Systola clava]	51	28	174
99	4	11357	4040	gi 1642795	TFIID subunit TAF155 [Homo sapiens]	51	25	684
109	1	2852	1428	gi 158920	rood (real) polypeptide (NA 1-673) [Bacillus subtilis] pir S0648 S0648	51	27	1425
					probable zod protein - Bacillus subtilis sp P13484 TAGE_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPSA-GLUCOSYLTRANSFERASE (EC 2.4.1.51) (TECHOCIC ACID BIOSYNTHESIS PROTEIN B).			
109	9	6007	6693	gi 111204815	hypothetical protein [SP:PI2662] [Haemophilus influenzae]	51	23	687
112	13	1066	2152	pir S05330 S053	maltose-binding protein precursor - Enterobacter aerogenes	51	42	1287
112	13	14432	112855	gi 1408857	YenU [Escherichia coli]	51	29	1578
114	9	9725	6867	gi 1435098	orf1 [Mycoplasma capricolum]	51	30	759
115	1	1	912	gi 1143110	UNP Y1005w [Succinatomyces cutivulvula]	51	25	912
127	10	9647	10477	gi 1204314	H. influenzae predicted coding region H10056 [Haemophilus influenzae]	51	37	831
152	9	6814	7356	gi 1431929	Hun1 regulatory protein [Mycoplasma sp. 1]	51	38	543
154	2	575	1153	gi 1227044	unknown [Mycobacterium tuberculosis]	51	36	579
154	7	6587	5634	gi 1409286	lwtU [Bacillus subtilis]	51	27	954
171	8	6943	6236	gi 1205484	Hypothetical protein (SP:PI3918) [Haemophilus influenzae]	51	32	708
184	1	1	291	gi 1466886	BR1996_C1206 [Mycobacterium leprae]	51	33	291
212	5	1501	2139	pir A5605 A456	nature-parasite-infected erythrocyte surface antigen MESA - Plasmodium falciparum	51	23	639
228	2	707	1378	gi 18204	inuclear protein [Protophyla melanogaster]	51	27	672
236	8	8137	7481	gi 149272	Asparagine (Bacillus licheniformis)	51	31	657
243	4	4637	3546	gi 1511102	melavionate kinase [Methanococcus jannaschii]	51	29	1092

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
10								
257	4	350	3373	[gi 1 204579]	[H. influenzae predicted coding region H10326 [Haemophilus influenzae]	51	22	168
258	3	2397	1609	[gi 1 60299]	[glutamic acid-rich protein [Plasmodium falciparum] PIR/AS514 AS4514]	51	34	789
265	5	2419	3591	[gi 1 80841]	[glutamic acid-rich protein precursor - <i>Plasmodium vivax</i>]	51	32	1173
288	2	518	748	[gi 1 336162]	[SCPB [Streptococcus agalactiae]	51	34	231
316	9	5817	7059	[gi 1 3953]	[lpa-29d gene product [Bacillus subtilis]	51	39	1233
332	2	3775	2057	[gi 1 203012]	[lmcS [Thermus aquaticus thermophilus]	51	26	1719
364	4	3816	4991	[gi 1 528991]	[unknown [Bacillus subtilis]	51	32	1176
440	2	448	684	[gi 1 2819]	[transf erase (GalT) (AA 1 - 687) [Kluyveromyces lactis] P1501407 XUVKG	51	32	217
					[UDPGlucose 4-epimerase (EC 5.1.3.2) - Yeast <i>yeastomyces maritimus</i> var. lactis]			
495	2	1353	1177	[gi 1 297661]	[protease C [Erwinia chrysanthemi]	51	41	177
495	3	2287	1718	[gi 1 5133317]	[serine rich protein [Entamoeba histolytica]	51	25	570
506	1	840	1421	[gi 1 455220]	[cII protein [Bacteriophage P1]	51	33	420
600	1	1474	983	[gi 1 587532]	[orf, len: 201, CAI: 0.16 [Saccharomyces cerevisiae] PIR/S48818 S48818]	51	30	492
					[hypothetical protein - yeast (Saccharomyces cerevisiae)]			
607	3	479	934	[gi 1 511524]	[hypothetical protein (SP_P137002) [Methanococcus jannaschii]	51	40	456
686	2	127	600	[gi 1 93017]	[endothelitis specific antigen [Enterococcus faecalis]]	51	30	474
726	1	31	230	[gi 1 351851]	[unknown [Prochlorococcus marinus]	51	45	198
861	1	176	652	[gi 1 0145]	[dehydroquininate dehydratase [Bacillus subtilis]]	51	34	477
869	1	782	393	[gi 1 0100]	[rodC (tag) polypeptide (AA 1-146) [Bacillus subtilis] P150049 S06049]	51	23	390
					[rodC protein - <i>Bacillus subtilis</i> P151485 TAGF_2BGSU TECHNIC ACID BIOSYNTHESIS PROTEIN F]			
1003	1	642	322	[gi 1 279707]	[hypothetical phosphoglycerate mutase [Saccharomyces cerevisiae]]	51	39	321
1046	2	866	624	[gi 1 01057]	[glycosyltransferase [Escherichia coli]]	51	29	243
1467	1	703	352	[gi 1 51175]	[H. Jannaschii predicted coding region HJ1177 (Methanococcus jannaschii)]	51	32	351
2558	1	457	230	[sp P10582 DROM_]	[DNA POLYMERASE (EC 2.7.7.7) (S-1 DNA ORF 3)]	51	26	228
3003	1	779	399	[gi 1 09543]	[CbcC protein [Erwinia chrysanthemi]]	51	27	381
3604	1	1	399	[pnu JC210 JC42]	[3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) - mouse]	51	37	199
3732	1	2	316	[gi 1 459206]	[acyl-CoA synthetase [Escherichia coli]]	51	33	315

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start	Stop	match accession	match gene name	% sim	% Ident	length (nt)	
10	10	[ID (nt)	[nt)						
3791	1	1	2	274	gi 1061351				
3995	1	46	336	gi 16346	[semaphorin III family homolog [Homo sapiens]	51	37	273	
					[putative synthetase [Bacillus subtilis]				
4193	1	612	307	gi 142749	[ribosomal protein L12 (AA 1-179) [Escherichia coli] [gi 504776 XXEPL peptide N-acetyltransferase C-terminal (EC 2.3.1.-) - <i>Escherichia coli</i>]	51	25	306	
4539	1	167	185	gi 1408494	[homologous to penicillin acylase [Bacillus subtilis]	51	40	183	
4562	1	442	239	gi 1456280	[coded for by <i>C. elegans</i> cDNA clone?; Similar to hydroxymethyl glutaryl-CoA synthase [Caenorhabditis elegans]	51	35	204	
1	4	3576	4859	gi 1559160	[GRAIL score: null; cap site and late promoter motifs present; putative [Autographa californica nuclear polyhedrosis virus]	50	44	1284	
11	7	4044	5165	gi 11146207	[putative [Bacillus subtilis]]	50	35	1122	
11	13	10509	9496	gi 1208451	[hypothetical protein [Synecchocystis sp.]]	50	39	1014	
19	1	2014	1018	gi 113966	[ipa-42d gene product [Bacillus subtilis]]	50	29	1017	
20	111	8516	8407	gi 1323159	[ORF YOR103W [Saccharomyces cerevisiae]]	50	28	180	
24	5	5408	4924	gi 1496280	[structural protein [Bacteriophage Tuc2005]]	50	29	505	
34	4	1926	2759	gi 11031966	[Yq10 [Bacillus subtilis]]	50	36	814	
36	30	122665	123440	gi 11072179	[Similar to dihydroflavonol-4-reductase [maize, petunia, tomato] [Caenorhabditis elegans]]	50	32	576	
47	2	1705	2376	gi 115015	[F6A protein [Staphylococcus aureus]]	50	29	1272	
56	13	15290	15841	gi 1606096	[ORF-1167; end overlaps end of 01 by 14 bases; start overlaps f174, ther starts possible [Escherichia coli]]	50	30	552	
57	1	2135	1077	gi 1640922	[xyitol dehydrogenase [unidentified <i>homozygote</i>]]	50	29	1059	
58	2	678	1761	gi 143725	[putative [Bacillus subtilis]]	50	29	1134	
88	6	4393	3884	gi 1072179	[Similar to dihydroflavonol-4-reductase [maize, petunia, tomato] [Caenorhabditis elegans]]	50	32	510	
89	5	1700	3356	gi 1276658	[ORF214 gene product [Porphyra purpurea]]	50	25	345	
141	1	1	3	239	gi 1476024	[carboxyl phosphate synthase II [Plasmidum falciplarum]]	50	33	237
151	1	186	626	gi 1403441	[unknown [Mycobacterium tuberculosis]]	50	35	441	
166	7	11065	9623	gi 1895747	[putative cell operon regulator [Bacillus subtilis]]	50	32	1443	
201	6	5284	5096	gi 11160229	[circumsporozoite protein [Plasmodium reichenowi]]	50	42	189	
206	122	30784	129555	gi 1052754	[Larp integral membrane protein [Lactococcus lactis]]	50	24	1230	

TABLE 2

5
10
15
20
25
30
35
40
45
50
55

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	sim	ident	length (nt)
211	4	1523	1927	gi 41031	[ORF]7 [Bacillus subtilis]	50	29	405
214	4	2401	2295	sp P37348 YEEC_	[HYPOTHETICAL PROTEIN IN ASPS 5 REGION (FRAGMENT)]	50	37	885
228	7	5068	4406	gi 13580	[envelope protein (Human immunodeficiency virus type II) pir15355 S25835	50	35	663
					[envelope protein - human immunodeficiency virus type I (fragment) (SUB 1-77)]			
222	2	3048	1723	gi 1408485	[B63G gene product [Bacillus subtilis]]	50	22	1326
273	2	1616	984	gi 1198196	[phosphoglycerate mutase (Saccharomyces cerevisiae)]	50	28	631
328	12	2507	1605	gi 1148996	[lipoprotein [Haemophilus influenzae]]	50	26	903
332	4	5469	3802	gi 1536547	[DNA polymerase family X [Thermus aquaticus]]	50	27	1688
342	5	1473	3931	gi 145562	[leX-box binding factor (Bacillus discoidineum)]	50	35	459
352	1	1478	741	gi 1280101	[ORF2 gene product [Bacillus megaterium]]	50	29	738
408	7	5299	5523	gi 11365	[ORF2136 (Murchantia polymorpha)]	50	27	223
420	3	650	1825	gi 175842	[UDP-sugar hydrolase (Escherichia coli)]	50	30	1176
464	1	1	591	gi 487282	[Na ⁺ -ATPase subunit J [Enterococcus hirae]]	50	29	591
472	2	1418	864	gi 155875	[BglR [Lactococcus lactis]]	50	23	535
520	1	23	541	gi 1567036	[CapE (Staphylococcus aureus)]	50	27	519
529	1	6	410	gi 11256632	[25% identity to the E. coli regulatory protein HfqA: putative [Bacillus subtilis]]	50	34	405
534	5	7726	6059	gi 1291671	selected as a weak suppressor of a mutant of the subunit A ₂ B ₂ C of DNA dependent RNA polymerase I and III (Saccharomyces cerevisiae)]	50	18	1664
647	1	2990	1497	gi 1405568	[trial protein shares sequence similarity with a family of apolipoproteins (Plasmodium ps44)]	50	31	1494
664	3	1133	711	gi 410007	[leukocidin F component (Staphylococcus aureus, MRSA No. 4, Peptide, 23 aa)]	50	32	423
678	1	1	627	gi 128032	[EF (Streptococcus suis)]	50	29	627
755	3	947	1171	gi 150572	[cytochrome c1 precursor (EC 1.10.2.2) [Paracoccus denitrificans] gi 15465	50	37	225
					[cytochrome c1 (AA 1-450) [Paracoccus denitrificans] pir C29413 C29413 ubiquinol--cytochrome-c reductase (EC 1.10.2.2) [Yochrome c1 precursor - Paracoccus denitrificans] sp P13627 C1]			
827	1	1363	683	gi 142020	[heterocyst differentiation protein (Anabaena sp.)]	50	21	681
892	1	3	752	gi 145485	[B63G gene product [Bacillus subtilis]]	50	27	750
910	2	438	887	gi 1149727	[tyrosine-specific transport protein (Haemophilus influenzae)]	50	25	450

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start	Stop	match	match gene name	% sim	% ident	length (nt)
10	10	(nt)	(nt)	acquisition				
933	1	524	760	[gi 202451]	[cell division inhibitor (Haemophilus influenzae)]	50	32	237
973	1	424	236	[gi 386247]	[orf3 gene product (Saccharomyces cerevisiae)]	50	40	189
1009	1	653	429	[gi 153727]	[M protein (Group G streptococcus)]	50	28	225
1027	1	511	257	[gi 413934]	[lpa-10r gene product (Bacillus subtilis)]	50	25	255
1153	2	556	326	[gi 173676]	[hccA (Salaligenes xylosyoxydans)]	50	36	231
1222	1	798	400	[gi 140485]	[B65G gene product (Bacillus subtilis)]	50	21	399
1350	1	692	399	[gi 289272]	[ferrichrome-binding protein (Bacillus subtilis)]	50	32	294
2945	1	366	184	[gi 171704]	[hexapeanyl pyrophosphate synthetase (CQ01) (Saccharomyces cerevisiae)]	50	34	183
2968	2	1604	804	[gi 397326]	[clumping factor (Staphylococcus aureus)]	50	31	801
2998	2	657	394	[gi 495696]	[F54E7.3 gene product (Cenorhabditis elegans)]	50	40	264
3046	1	516	306	[pN151R19 S138]	[acyl carrier protein - Anabaena variabilis (fragment)]	50	32	201
3063	1	547	275	[gi 174390]	[lucA gene product (Escherichia coli)]	50	29	273
3174	1	3	146	[gi 151900]	[alcohol dehydrogenase (Rhodobacter sphaeroides)]	50	31	144
3792	1	625	314	[gi 1001423]	[hypothetical protein (Synchocystis sp. P1)]	50	35	312
3800	1	2	262	[gi 144713]	[NAD-dependent beta-hydroxybutyryl coenzyme A dehydrogenase Clostridium acetobutylicum]	50	28	261
3946	1	373	188	[gi 176765]	[cytochrome b (Myrmecia pilosula)]	50	38	146
3984	1	578	291	[apl P1748 YCE_]	[HYPOTHETICAL PROTEIN IN AAPS 5 REGION (FRAGMENT)]	50	37	288
3987	10	8250	7835	[gi 204367]	[hypothetical protein (GB 014003_278) (Haemophilus influenzae)]	49	30	366
46	16	113802	114848	[ccl 166660]	[acd: B108 F1_14 (Mycobacterium leprae)]	49	24	1047
59	5	2267	3601	[gi 66304]	[ORF_0462 (Escherichia coli)]	49	27	1335
112	18	117884	118615	[gi 559502]	[hd4 protein (AA 1 - 409) (Caenorhabditis elegans)]	49	25	732
118	9	6973	7902	[gi 103953]	[esterase (Acinetobacter calcoaceticus)]	49	29	930
217	6	4401	5138	[gi 496534]	[fibronectin/fibrinogen-binding protein (Streptococcus pyogenes)]	49	31	738
220	12	111803	112657	[gi 197526]	[clumping factor (Staphylococcus aureus)]	49	31	855
228	4	1842	12492	[pir 1522692 S2336]	[hypothetical protein 9 - Plasmodium falciparum]	49	24	651
268	1	5016	12614	[gi 143047]	[ORF8 (Bacillus subtilis)]	49	26	2401

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (ID)	Stop (Incl)	match accession	match gene name	% sim	% ident	length (int)	
271	1	2	1164	1373	gi 1001257	hypothetical protein [Synchocystis sp.]	49	18	210
300	1	3	4340	3160	gi 11510796	hypothetical protein [GP-X91006_2] [Methanococcus jannaschii]	49	26	1161
381	1	2281	1142	gi 139301	[matches PS00041: Bacterial regulatory proteins, arac family signature] [Escherichia coli]	49	29	1140	
466	1	3	947	gi 11303863	YQ9P [Bacillus subtilis]	49	26	945	
666	1	179	191	gi 633112	ORF1 [Streptococcus sobrinus]	49	29	189	
670	1	2	403	1014	gi 11122758	[unknown (Bacillus subtilis)]	49	32	612
709	1	1433	1	795	gi 1114830	xpC [Bacillus subtilis]	49	29	639
811	1	943	1	473	gi 1401786	[phosphomannomutase] [Mycoplasma pneumoniae]	49	29	471
1052	1	422	213	gi 1303799	[YQ9N (Bacillus subtilis)]	49	21	210	
1800	1	142	172	gi 216300	[peptidoglycan synthesis enzyme [Bacillus subtilis] sp P07595 WURG_BACSU]	49	28	171	
					MURG PROTEIN UDP-N-ACETYLGUCOSAMINE-N-ACETYLGLURAMYL-PENTAPEPTIDE PYROPHOSPHORYL-UNDECAPENOL N-ACETYLGLUCOSAMINE TRANSFERASE]				
2430	1	2	376	sp P22743 YFGA-	[HYPOTHETICAL 36.2 KD PROTEIN IN NDK-CPPE INTERGENIC REGION]	49	26	375	
3096	1	542	273	gi 1515360	[surfactin synthetase [Bacillus subtilis]]	49	25	270	
32	1	4	3771	gi 1100	hepatocyte nuclear factor 4 gamma [HMGgamma] [Homo sapiens]	48	36	672	
3K	1	1	609	gi 1205790	[H. influenzae predicted coding region H1155 [Haemophilus influenzae]]	48	28	609	
45	6	5021	6427	gi 1154267	[unknown [Mycobacterium tuberculosis]]	48	20	1407	
59	14	16346	31096	gi 1197316	[Lmp] protein [Mycoplasma hominis]	48	28	14751	
61	1	1	608	gi 1511555	[quinolone resistance nora protein protein [Methanococcus jannaschii]]	48	30	606	
61	1	3	3311	13646	YQ9U [Bacillus subtilis]	48	29	336	
114	1	98	415	gi 1671708	[suis] homology; similar to Drosophila melanogaster suppressor of able (tau(s)) protein, Swiss-Prot Accession Number P22233 [Drosophila virilis]	48	25	318	
121	1	1131	610	gi 11314584	[unknown (Sphingomonas SB8)]	48	29	522	
136	1	2014	1280	gi 11205968	H. influenzae predicted coding region H11738 [Haemophilus influenzae]	48	23	735	
171	10	8220	9557	gi 11208434	[hypothetical protein [Synchocystis sp.]	48	34	1338	
175	1	1625	1814	gi 1396400	similar to eukaryotic Na ⁺ /H ⁺ exchangers [Escherichia coli] sp P12701 YJCF_ECOLI HYPOERTICAL 60.5 KD PROTEIN IN SOXR-ACS NTERGENIC REGION (059).	48	29	1812	
194	1	2	385	gi 11510493	H. jannaschii predicted coding region HJ0419 [Methanococcus jannaschii]	48	25	384	

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF	Start	Stop	match	match gene name	% sim	% ident	length (nt)		
ID	ID	int	int	accession						
1	197	1	901	452	gi 1045714	spermidine/potresine transport ATP-binding protein [Mycoplasma genitalium]	1	48	25	450
203	1	1	396	gi 190288	protein localized in the nucleoli of pea nuclei; ORF; putative P1um	48	29	396		
204	1	1363	698	gi 529202	No definition line found [Caenorhabditis elegans]	48	25	666		
206	20	14815	27760	gi 511490	gramicidin S synthetase 2 [Bacillus brevis]	48	27	7056		
212	1	1	2	166	gi 125899	lncolisin [Xenopus laevis]	48	34	165	
220	10	12652	11426	gi 141073	lsecy protein [Lactococcus lactis]	48	23	1227		
243	1	6	6450	15491	gi 1184118	lmalonate kinase [Methanobacterium thermophilicum]	48	30	960	
264	4	5434	2308	gi 1015903	ORF YJ181c [Saccharomyces cerevisiae]	48	26	2127		
441	1	1532	768	gi 142863	replication initiation protein [Bacillus subtilis] pir B26580 B26580	48	23	765		
444	5	398	5298	gi 1165836	replication initiation protein - Bacillus ubtilis	48	24	1401		
484	2	388	1110	gi 146551	transmembrane protein (kpD) [Escherichia coli]	48	18	723		
542	1	1425	2000	gi 528969 S289	N-carbamoyl sarcosine amidohydrolase [C. J. S. 1.59] - Arthrobacter sp.	48	27	576		
566	1	1	109	gi 153490	tetracycycin C resistance and export protein [Streptomyces lauseiensis]	48	24	1017		
611	1	1	730	gi 1103507	lunknow [Schizosaccharomyces pombe]	48	18	729		
624	1	1255	665	gi 148859	ORF B [Clostridium perfringens]	48	26	591		
146	1	1014	508	gi 5137506	parathyosin [Dirofilaria immitis]	48	27	507		
1020	1	66	930	gi 1149916	magnesium and cobalt transport protein [Methanococcus jannaschii]	48	10	865		
1227	1	1	174	gi 1493730	lipoygenase [Pisum sativum]	48	15	174		
1266	1	1	405	gi 1882452	ORF (211; alternate name yga; orfs of XI4436 [Escherichia coli] gi 41425	48	24	405		
					ORF5 (AA 1-197) [Escherichia coli] (SUB 15-211)					
2071	1	707	381	gi 1408486	HS74A gene product [Bacillus subtilis]	48	25	327		
2398	1	1463	233	gi 1500401	reverse gyrase [Methanococcus jannaschii]	48	40	231		
2425	1	476	246	gi K1H853 H485	G1 protein - poxvirus (strain HP44) (fragment)	48	40	231		
2432	1	446	225	gi 1351703	trio (Homo sapiens)	48	33	222		
2453	1	794	399	gi 142850	division initiation protein [Bacillus subtilis]	48	29	396		
2998	1	459	236	gi 577569	PepV [Lactobacillus delbrueckii]	48	31	234		

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4027	2	492	301	[pir:SS1171SS511]	[trans-activator protein - Equine infectious anemia virus]	48	32	197
4	12	3641	2232	[gi 1103989]	[Yohi (Bacillus subtilis)]	47	24	1410
24	12	599	1084	[gi 5100083]	[PCP-1 gene product (Bradyia hygida)]	47	28	496
43	2	196	1884	[gi 1103055]	[esterase (Acinetobacter lwoffii)]	47	26	600
44	22	16118	15108	[gi 1151555]	[Unknown [Mycobacterium tuberculosis]]	47	27	1689
69	7	7141	6710	[gi 438466]	[quinolone resistance mva protein protein in [Methanococcus jannaschii]]	47	31	1011
81	4	5022	4279	[gi 4466862]	[possible operon with orfG. Hydrophobic, no homologue in the database; putative (Bacillus subtilis)]	47	29	412
120	12	9115	8863	[gi 1927340]	[psb1: Bl496 C2.189 (Mycobacterium leprae)]	47	24	744
142	1	2022	1174	[gi 486143]	[D50509-27p; CAY: 0.12 (Saccharomyces cerevisiae)]	47	38	273
168	1	2178	1093	[gi 1177554]	[hypothetical Ec8B protein (Bacillus subtilis)]	47	32	849
263	1	1864	943	[gi 1112822]	[D-alanine racemase cds (Bacillus subtilis)]	47	29	1086
279	1	1109	561	[gi 516608]	[2 predicted membrane helices, homology with <i>B. subtilis</i> men Orf3 Rowland et al. unpublished accession number M741831, approximately 1 minute on updated Rudd map; putative [Escherichia coli] sp P37355 YFB_ECOLI HYPOPTICAL 26.7 KO PROTEIN IN MEND+ENB]	47	34	942
345	2	2620	1676	[gi 1204635]	[hypoprotease (Haemophilus influenzae)]	47	31	549
389	2	152	400	[gi 456562]	[G-box binding factor (Dictyostelium discoideum)]	47	28	945
391	1	1	831	[gi 1420556]	[myo-inositol transporter (Schizosaccharomyces pombe)]	47	32	249
404	3	2022	2773	[gi 1255425]	[C3G8-2 gene product (Caenorhabditis elegans)]	47	19	831
529	5	2145	3107	[gi 1103973]	[Y01V (Bacillus subtilis)]	47	17	702
565	2	2321	1257	[gi 142874]	[processing protease (Bacillus subtilis)]	47	29	963
654	1	962	483	[gi 1243353]	[orf 5' of ECRP3 (herpesvirus saimiri IVS, host-squirrel monkey, epide, 407)]	47	28	1065
692	1	115	633	[gi 1150756]	[40 kDa protein (Plasmid pM)]	47	25	519
765	1	1634	819	[gi 1256621]	[26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative (Bacillus subtilis)]	47	28	816

TABLE 2

5
10
15
20
25
30
35
40
45
50

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
R25	1	211	1033	[gi 1197326]	[clumping factor (Staphylococcus aureus)]	47	32	813
914	1	1	615	[gi 158073]	[leucyltRNA synthetase (Plasmodium falciparum)]	47	29	615
1076	1	1	753	[gi 1147557]	[leucyltRNA synthetase (Bacillus circulans)]	47	33	753
1351	1	793	398	[gi 1755153]	[ATP-binding protein (Bacillus subtilis)]	47	20	396
4192	1	3	293	[gi 145636]	[putative (Escherichia coli)]	47	24	291
5	6	4708	4361	[gi 105080]	[laminin heavy chain (Entamoeba histolytica)]	46	30	348
11	4	2771	3058	[gi 1036169]	[Yel40p (Saccharomyces cerevisiae)]	46	28	282
46	11	10518	110300	[gi 1246901]	[ATP-dependent DNA ligase (Candida albicans)]	46	28	219
61	4	3941	7930	[gi 198032]	[EF IS (Streptococcus suis)]	46	35	1990
132	4	5028	4093	[gi 111057]	[hypothetical protein SPAP4869 (Methanococcus jannaschii)]	46	25	936
170	4	4719	3622	[gi 151910 5S19]	[64 protein - <i>Sauvoleshmania tarantoiae</i>]	46	26	1068
191	7	9543	8284	[gi 1041334]	[P5405.7 (Caenorhabditis elegans)]	46	25	1260
253	1	1	396	[gi 120449]	[dihydrofolate acetyltransferase (Haemophilus influenzae)]	46	35	396
264	3	437	973	[gi 180189]	[cerebellar-degeneration-related antigen (CDR34) (Homo sapiens) gi 1182737]	46	29	517
					[cerebellar degeneration-associated protein (Homosapiens) pir A29770 A29770 cerebellar degeneration-related protein - human]			
273	1	485	1285	[gi 1607573]	[envelope glycoprotein C2V3 region (Human immunodeficiency virus type 1)]	46	35	201
350	1	1	563	[gi 137052]	[ORF f216 (Escherichia coli)]	46	35	561
JH4	1	1	2	[gi 121884]	[lamininolyase (Haemophilus influenzae)]	46	31	861
410	4	1876	12490	[gi 1110518]	[proton antiporter efflux pump (Mycobacterium smegmatis)]	46	24	615
432	1	2663	1455	[gi 1197634]	[orf1; putative transporter; Method: conceptual translation supplied by author (Mycobacterium smegmatis)]	46	27	1209
458	1	2419	1211	[gi 15470]	[putative protein (Bacteriophage SP1)]	46	30	1209
517	5	2477	4192	[gi 1523812]	[orf5 (Bacteriophage A2)]	46	23	1716
540	3	1512	1285	[gi 1215615]	[pacA (Bacteriophage P1)]	46	30	228
587	2	649	1242	[gi 537148]	[ORF_f181 (Escherichia coli)]	46	29	594
1218	1	747	391	[gi 1203456]	[single-stranded-DNA-specific exonuclease (Haemophilus influenzae)]	46	30	357

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3685	1	1	402	gi 450688	hsdM gene of Ecopri gene product (Escherichia coli) pif S18437 S18437 hsdM protein - Escherichia coli pif S09629 S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	46	33	402
4176	1	673	338	gi 151460	[F1H-C-1 gene product (Xenopus laevis)]	46	31	336
37	7	4813	5922	gi 606064	[ORF f408 (Escherichia coli)]	45	24	1110
38	16	11659	12004	gi 452192	[protein Tyrosine phosphatase (PTP-BAS type 2) (Homo sapiens)]	45	24	306
87	2	1748	12407	gi 1064813	[homologous to sp:PHOR_BAC01 (Bacillus subtilis)]	45	23	660
103	12	14382	113305	gi 1101307	[hypothetical protein (Synochocystis sp.)]	45	22	798
112	14	14791	13811	gi 1120189	[H. influenzae predicted coding region H10131 (Haemophilus influenzae)]	45	23	981
145	4	4483	3462	gi 1220578	[open reading frame (Hus musculus)]	45	20	1023
170	6	6329	4965	gi 1238657	[apocytochrome c oxidase, subunit 1 homolog (Escherichia coli, K12, <i>l</i> epid.) 514 aa]	45	27	1365
206	2	5210	4346	gi 122056	[laminotransferrase (Haemophilus influenzae)]	45	27	885
228	1	60	716	gi 160299	[glutamic acid-rich protein (Plasmodium falciparum) pif A54514 A54514	45	23	657
288	1	2	1015	gi 11255425	[glutamic acid-rich protein precursor - Plasmodium aliciparum]	45	23	1014
313	3	43159	3128	gi 1581140	[NADH dehydrogenase (Escherichia coli)]	45	30	1212
332	1	914	459	gi 1820566	[F71A4.2 (Caenorhabditis elegans)]	45	20	456
344	1	3	221	gi 1171225	[kininsin-related protein (Caenorhabditis cerevisiae)]	45	26	219
441	2	1501	1073	gi 142863	[replication initiation protein (Bacillus subtilis) pif B26580 B26580]	45	27	429
672	1	2	962	gi 11511334	[1H. Jannaschii predicted coding region H01133 (Methanococcus jannaschii)]	45	22	981
763	3	1345	851	gi 606180	[ORF f510 (Escherichia coli)]	45	24	495
886	3	379	846	gi 726426	{similar to protein kinases and C. elegans proteins F3C12.8 and J7C12.5 [Caenorhabditis elegans]}	45	30	468
948	1	3	473	gi 156400	{tyrosin heavy chain (Iggyne unc-54) [Caenorhabditis elegans] [pif A3J958 HMMW myosin heavy chain B - Caenorhabditis elegans] [sp P02556 HNSB CAEL KYOSIN HEAVY CHAIN B (HNC B)}	45	25	471
1158	1	2	376	gi 441155	[transmission-blocking target antigen (Plasmodium falciparum)]	45	35	375
2551	1	4	285	gi 1276705	[ORF287 gene product (Porphyra purpurea)]	45	28	282
3967	1	42	374	gi 976025	[Hsa [Escherichia coli]]	45	28	333

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

5
10
15
20
25
30
35
40
45
50
55

Contig	ORF	Start	Stop	match	match gene name	% sim	% ident	length	
ID	ID	(nt)	(nt)	accession		(nt)	(nt)	(nt)	
52	1	6931	5846	gi 4673728	unknown [Bacillus subtilis]	44	22	1086	
138	8	6475	6849	gi 173028	thioredoxin II [Saccharomyces cerevisiae]	44	28	375	
221	5	7032	5617	gi 153490	tetracycline C resistance and export protein [Streptomyces laevisiens]	44	21	1416	
252	2	1331	1122	gi 1204989	hypothetical protein [GBU00022_9] [Haemophilus influenzae]	44	30	210	
263	1	2	3265	gi 11136221	carboxypeptidase [Sulfolobus solfataricus]	44	26	1173	
365	4	4963	3524	gi 1296822	lorti gene product [Lactobacillus helveticus]	44	31	1440	
543	3	1315	1833	gi 1083250	low homology to P20 protein of <i>Bacillus licheniformis</i> and bleomycin acetyltransferase of <i>Streptomyces verticillatus</i> [Bacillus subtilis]	44	24	519	
544	4	1	3942	4892	gi 951460	IFIM-C gene product [Xenopus laevis]	44	32	951
792	1	1	1224	613	gi 1205680	high molecular weight neurofilament [Rattus norvegicus]	44	28	612
44	18	111303	11911	gi 1511614	methylbacteriopheophytin-guanine dinucleotide biosynthesis protein A [Methanococcus jannaschii]	43	27	609	
59	8	1	3665	5128	gi 153490	tetracycline C resistance and export protein [Streptomyces laevisiens]	43	21	1464
59	110	1	5516	7527	gi 153022	lipase [Staphylococcus epidermidis]	43	22	1992
99	1	1	1346	681	gi 11419051	unknown [Mycobacterium tuberculosis]	43	21	666
110	8	1	9402	11234	gi 397526	clumping factor [Staphylococcus aureus]	43	21	2733
412	3	2782	2303	gi 16054014605	leporoocyte surface protein 2 - Plasmodium yoelii (fragment)	43	29	480	
519	3	2547	3122	gi 00653010HSU_	SULFIDE DEHYDROGENASE (FLAVOCYTOCHROME C) FLAVOPROTEIN CHAIN PRECURSOR [EC 1.8.2.-] (FC) (FESD).	43	21	576	
4	13	12053	11321	gi 1295671	selected as a weak suppressor of a mutant of the subunit AC40 of DNA endonuclease I and III [Saccharomyces cerevisiae]	42	18	1269	
94	2	1768	1091	gi 501027	ORF2 [Trypanosoma brucei]	42	31	678	
127	4	5791	4550	gi 12025	ORF1 gene product [Escherichia coli]	42	21	1242	
297	1	3	1515	1016	gi 142790	ORF1, putative [Bacillus firmus]	42	25	480
144	6	1	4097	1525	gi 140320	ORF 2 [AA 1-203] (Bacillus thuringiensis)	42	30	573
512	1	1	2167	1115	gi 405937	l-type [Escherichia coli]	42	23	1053
631	1	2434	1223	gi 580920	probable robo protein - Bacillus subtilis sp P13847 [BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPHA-GLUCOSYCTANSE (EC 2.4.1.32) (PHOCHOLIC ACID BIOSYNTHESIS PROTEIN E).	42	24	1212	

TABLE 2

5
10
15
20
25
30
35
40
45
50*S. aureus* - Putative coding regions of novel proteins similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
86	2	1375	884	gi 109506	sporidine/spermine N1-acetyltransferase [Mus musculus] Dif S1430 S84340	42	19	621
191	12	14797	11075	gi 11103784	sporidine/spermine N1-acetyltransferase [Mus musculus] Dif S1430 S84340	42	25	393
4132	1	787	395	gi 1102310	protein tyrosine phosphatase [Dictyostelium discoideum] ORF4	42	25	393
212	1	2150	3127	gi 115873	observed 35.2kd protein [Mycobacteriophage 151] ORF4	41	22	723
213	1	1261	2000	gi 1633692	[Yersinia enterocolitica] ORF4; putative product: Method: conceptual translation supplied by author [Mycobacterium smegmatis]	41	22	723
408	4	2625	3386	gi 1197634	ORF4; putative product: Method: conceptual translation supplied by author [Mycobacterium smegmatis]	41	26	978
542	1	3	1103	gi 1457146	trichoptery protein [Plasmodium yoelii]	41	18	738
924	1	2	475	gi 1301481JH01	nucleolin - rat	41	10	474
1562	1	1	402	gi 1552184	asparagine-rich antigen Pfa35-2 [Plasmodium falciparum] Dif S2182 S27826	40	20	402
2385	1	518	261	gi 152251 S422	hypothetical protein 5 - fowlpox virus	40	18	256
4077	1	3	305	gi 1035055	coded for by <i>C. elegans</i> cDNA yk37g.5; coded for by <i>C. elegans</i> cDNA yk59.5; alternatively spliced form of F52C9.8b [Caenorhabditis elegans]	39	21	103
918	1	1003	503	gi 11255425	C13G8.2 gene product [Caenorhabditis elegans]	37	25	501
59	112	8294	106316	gi 153260	ISTAMP antigen [Plasmodium reichenowi]	36	24	2143
61	5	1550	8079	gi 129032	IF [Streptococcus suis]	36	19	4530
544	1	2507	3601	gi 11015903	ORF YJR15C [Saccharomyces cerevisiae]	35	22	1095
63	4	1949	3574	gi 1552195	circumsporozoite protein [Plasmodium falciparum] sp1P05691 CSP_PLAFL	32	27	1626
					CIRCUOSPOROZOITE PROTEIN (CS) (FRAGMENT)			

TABLE 2

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. entericus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)
4	1	1234	692
4	3	1712	2278
4	4	3703	1032
4	114	13073	112585
5	2	2539	1601
5	3	1532	1771
5	5	7	4741
5	9	7939	6422
5	112	6711	8547
6	4	2359	1982
6	1	349	176
11	8	5144	5983
11	9	5968	6498
11	10	6472	6284
11	16	10954	11271
12	5	5352	4942
12	6	4596	4862
15	3	1895	1650
16	10	11263	10835
18	2	1093	917
20	13	12158	10770
23	1	674	319
23	6	6118	5485
23	8	6376	5942

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF ID	Start ID	Stop ID	Start (nt)	Stop (nt)
23	9	7	651	6881	
23	15	112	618	112830	
24	4	4556	4185		
24	6	562	5241		
25	2	1824	2402		
31	2	505	849		
31	3	1177	1524		
31	4	2454	3005		
32	2	1765	1388		
32	9	7922	8575		
32	10	8591	8728		
32	11	9738	9379		
32	12	110797	110087		
34	2	1315	1049		
36	7	5226	5801		
36	11	7575	7261		
36	12	7424	7821		
37	4	3158	2964		
38	2	1585	980		
38	11	6425	6868		
38	12	116982	116371		
38	126	120253	120804		
38	127	140722	14164		
39	1	1	627		
40	1	805	404		
43	1	796	428		
44	4	12674	12324		

5

10

15

20

25

30

35

40

45

50

55

S. aurous - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)
44	5	2484	3263
44	14	110587	110129
44	120	113724	113536
44	121	113596	113994
45	7	16575	6297
46	8	6105	6520
46	112	110449	110916
46	117	115032	115424
47	1	286	1079
48	9	7620	7778
50	1	1612	962
50	2	1621	1316
51	1	738	370
51	5	2520	2245
53	1	442	287
53	7	6705	6319
54	7	9014	8709
55	1	592	326
55	3	1551	1228
56	1	1032	786
56	1	1	261
56	4	1970	1560
56	17	119092	118712
57	4	3694	3521
57	8	5456	5822
58	9	8895	8553
59	3	1346	1509

TABLE 3

TABLE 3

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
Id	Id	(nt)	(nt)
59	6	3026	2802
59	7	3770	3570
59	9	496	4563
59	111	7518	8378
59	113	110401	116403
62	1,2	2666	1521
62	111	5440	5757
63	1,1	1	336
67	1,1	900	1781
67	2	1774	2610
67	3	2591	3904
67	8	7110	6955
68	1	78	326
70	6	6761	5199
70	111	8935	8665
70	3	1590	1192
79	2	1509	1228
79	3	1411	1791
83	1	2	403
85	9	8300	8653
85	110	8969	8781
86	3	1426	1232
87	8	9187	9366
88	3	1620	1922
89	1	3	161
89	7	5042	4878
91	1	1098	550

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Coding	ORF	Start	Stop
ID	ID	(nt)	(nt)
91	3	398	3141
92	2	449	928
92	3	1938	1467
92	9	5638	6024
94	1	661	332
94	3	2445	1813
94	4	2503	2197
96	11	110601	11050
99	6	472	4533
99	7	5014	4784
100	8	7658	7287
102	7	4597	4368
101	1	2496	2035
104	1	2	694
104	2	699	1277
105	1	1235	693
105	3	3233	2655
105	1	3	221
106	3	1209	1155
107	1	1081	542
109	4	4025	3651
109	13	11625	11396
109	14	11981	112268
109	20	11701	117688
110	1	2	760
114	10	8764	9384
116	1	1	109

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. enteritidis - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop	
ID	ID	(nt)	(nt)	
116	1	3	623	4462
116	1	8	11039	9976
116	9	10113	10158	
120	1	5	3703	3320
120	6	4270	3869	
120	13	920	9844	
121	2	417	569	
126	1	3	1030	818
127	1	3	2688	3196
127	5	4084	4395	
131	6	6773	6438	
132	2	715	1655	
134	1	2	667	
135	2	512	258	
135	3	1124	729	
138	1	3	152	
138	7	6006	6463	
140	1	2040	1032	
140	2	2019	1513	
140	5	2387	2743	
142	2	1160	1288	
142	7	8830	7586	
143	7	729	6502	
144	1	1227	640	
146	1	2	511	
146	3	502	1350	
146	4	3673	2540	

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)
146	5	2874	3071
147	1	1	119
149	111	3956	3615
149	112	4036	3785
149	113	4507	4145
149	115	4807	4610
149	116	5495	5049
149	118	5739	5491
149	121	7416	7034
149	123	9216	8521
149	124	9681	9106
149	125	110639	9897
150	2	2303	1587
154	3	1795	1508
154	8	6506	6398
154	14	112704	112147
154	15	113531	112803
156	1	315	593
157	3	11033	12232
158	2	1471	1064
159	3	452	808
161	2	876	1808
161	6	4653	4279
161	7	4803	4540
161	8	4636	4717
161	11	5617	5618
163	2	1604	840

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins				
Contig	ORF	Start	Stop	
16	ID	(nt)	(nt)	
163	1	5	2996	234
161	7	2952	2647	
163	9	4905	5132	
164	3	1338	1147	
166	3	5213	4854	
168	4	2500	2068	
168	5	395	4158	
170	3	2317	2777	
171	2	2277	1450	
171	111	12576	11125	
172	1	3	278	
172	2	1940	1149	
173	1	1289	708	
173	5	7001	6114	
174	2	593	1105	
175	3	2552	2890	
175	5	3820	3335	
175	7	4342	4506	
182	4	5477	4986	
184	5	6043	5702	
188	2	1210	1755	
188	4	2637	2994	
189	6	2614	3039	
190	3	1988	2584	
191	1	1	153	
191	2	950	1659	
191	10	11176	113019	

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
td	ID	(nt)	(nt)
191	111	12902	122163
192	1	91	426
195	3	2306	1932
195	5	2899	2606
198	2	1016	1591
201	1	170	625
203	2	1781	1466
206	6	8930	7815
206	112	13947	13616
206	121	128068	127960
212	2	170	817
212	3	796	1167
212	7	3128	13436
212	9	3749	4075
213	1	1	705
214	2	1076	510
214	6	4064	3738
214	9	6500	6395
214	110	7064	7469
217	1	1927	965
218	1	178	637
218	3	1776	2156
220	2	1851	1369
220	3	3251	2262
220	7	8275	7208
220	8	10344	8661
220	9	11796	10216

TABLE 3

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start (nt)	Stop (nt)
226	1	3	659
226	1	2	2196
226	1	3	1459
226	1	3	1476
226	1	9	11426
227	1	1	10757
227	1	2	487
227	1	2	460
227	1	2	975
227	1	4	1835
227	1	5	2121
227	1	5	2052
227	1	5	2345
227	6	4760	3768
227	9	5591	6367
228	1	5	2503
228	1	5	2877
228	6	2846	3526
233	7	1	1944
233	7	1	3762
236	1	2	809
236	1	2	579
238	1	2	1975
238	1	2	1391
239	1	2	1417
239	1	2	905
241	1	5	4495
241	1	5	4334
242	1	2	1677
242	1	2	1363
243	1	1	127
243	1	1	576
244	1	1	1291
244	1	1	647
244	1	2	3035
244	1	2	1962
245	1	2	1614
245	1	2	1258
246	1	1	69
246	1	1	215
246	4	1	738
246	4	1	1713
249	3	1	1906
249	3	1	3712
250	1	1	494
250	1	1	249

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. AUROUS - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start 10 (nt)	Stop (nt)
254	1	1	156
256	1,2	956	1144
257	3	3700	3227
260	4	4906	4880
261	4	2196	2606
261	6	3214	3681
264	2	155	439
264	5	5252	4533
264	6	4739	5107
267	2	1323	931
268	4	510	4700
272	1	862	446
272	3	1200	1419
272	9	4691	4909
272	10	6469	6035
276	4	1746	1901
278	1	224	553
278	5	3299	3448
278	7	4849	5127
285	2	551	716
288	3	1756	1950
288	5	2055	2276
289	1	2107	1055
290	2	2234	1932
291	2	332	622
291	5	155	2051
295	3	1666	1349

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)
295	4	2278	2141
295	5	2220	2762
297	2	788	465
298	1	2	205
300	2	2300	1928
301	7	2794	2624
304	1	3	194
306	1	109	654
306	5	4036	4257
307	1	674	339
307	6	3665	3995
308	1	1	654
308	2	1120	599
308	4	2643	2332
313	2	2314	1919
314	1	10	702
316	2	982	1341
316	6	2798	3165
317	1	2	1114
317	3	4570	3458
321	6	5645	5217
321	7	6319	6140
321	8	7450	6794
322	2	827	543
326	2	165	1112
326	3	1117	1467
328	1	956	469

TABLE 3

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aurous - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start (nt)	Stop (nt)
1	1	1	1
328	1	5	3452
329	1	3	719
329	2	2	783
329	3	1	1471
330	1	1	576
330	2	1	1447
332	1	3	2353
332	2	7	4971
333	1	2	3295
335	1	1	864
337	1	2	95
340	1	2	1658
341	1	1	3
341	2	1	2476
341	3	1	3618
341	4	1	3929
344	5	1	3197
345	1	1	1532
346	2	1	221
350	3	1	1410
352	1	2	2178
352	2	1	7316
352	3	7	7967
352	4	6	8906
352	5	9	10171
359	1	1	1
362	1	1	3

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Accession ID	ORF ID	Start (nt)	Stop (nt)
364	2	2158	1808
364	8	11094	110714
365	2	1612	1313
365	5	4610	4090
365	7	4380	6239
366	1	520	1719
367	3	906	1085
368	1	748	494
375	1	2	136
380	3	1351	1097
389	1	1	276
390	1	2	877
390	2	1373	1349
391	2	751	560
395	1	391	197
396	1	2132	1068
398	3	1344	1141
399	1	176	669
401	3	566	847
402	2	100	465
404	8	561	5370
408	2	13507	2269
408	3	2875	2672
408	5	1524	4423
410	3	2111	1890
413	1	860	488
416	1	607	320

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. enterica - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
ID	ID	(nt)	(nt)
416	2	578	847
416	3	2195	1520
417	1	3	179
417	2	161	616
420	2	788	513
422	2	357	677
431	2	856	1407
432	2	446	1084
433	1	1	417
433	3	2311	2033
434	1	942	535
434	2	2089	1225
440	1	1	450
442	2	1289	3320
443	3	1873	1520
444	1	1	696
444	7	6761	6366
451	1	940	614
453	2	896	636
453	8	383	4786
455	1	434	219
455	2	472	930
459	1	265	687
462	1	2	247
466	2	1494	907

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start (NP)	Stop (ID)	Start (In)	Stop (In)
467	1	654	349		
468	1	2	250		
469	1	1488	925		
469	3	2366	3372		
469	4	3464	3706		
470	1	77	538		
470	6	4098	3694		
470	7	6310	5686		
470	9	7351	8181		
470	110	8175	9773		
471	1	940	500		
471	2	1562	1017		
476	1	70	267		
477	1	2	760		
477	3	1764	2081		
477	4	2066	2332		
480	5	4016	4261		
481	2	956	480		
486	3	613	774		
487	6	1795	2112		
488	1	713	359		
492	1	127	675		
493	1	2	520		
493	2	496	1242		
502	3	1149	1571		
504	1	650	346		
505	5	4566	4150		

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative ending regions of novel proteins not similar to known proteins

Centig	ORF ID	Start (nt)	Stop (nt)
511	2	1701	1232
512	2	583	747
515	1	609	812
517	4	2179	2511
520	4	2097	2360
520	6	3908	3669
527	1	1	498
528	1	637	315
529	2	1679	1104
530	7	5298	5534
536	1	308	156
538	1	1302	736
538	3	2203	2880
538	5	3531	3121
538	6	4368	3731
540	1	956	664
540	2	1405	1031
541	1	69	433
541	2	719	4322
542	2	1048	1272
545	2	1012	734
551	1	2145	1129
555	2	892	704
558	3	1337	1154
558	4	1760	1458
558	5	2105	1821
558	6	2166	2020

TABLE 3

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
	1b	1D	Int
558	7	2336	2322
558	8	3053	2802
571	1	108	156
571	3	994	1206
577	1	1	2
577	2	163	453
579	1	1	477
579	2	4784	1200
583	1	1988	996
585	1	946	539
587	1	22	573
586	2	1496	1372
588	3	1742	1554
590	1	47	334
592	2	1455	1141
593	1	1	775
593	2	617	1122
595	1	87	890
596	3	1593	1435
602	1	8	169
603	5	1071	1469
606	1	322	768
607	5	1444	13226
610	1	1039	541

TABLE 3

10

15

20

25

30

35

40

45

50

55

S. aurous - Putative coding regions of novel proteins not similar to known proteins

-coding		orf	Start	Stop
Pb	Id	Int	Int	Int
612	1	3	500	
616	1	991	650	
617	2	736	491	
622	1	36	347	
625	4	2046	2569	
627	1	67	210	
628	1	901	452	
631	3	4789	4004	
634	1	1488	759	
636	1	189	368	
636	2	1929	1063	
637	2	2323	1984	
638	1	227	1081	
639	1	510	261	
639	2	1377	811	
641	1	118	444	
642	3	1615	1331	
642	4	2260	1847	
643	1	3	608	
645	4	1534	1758	
645	6	2075	2321	
645	7	2900	2488	
648	1	2	1005	
660	1	77	601	
660	2	516	872	
661	1	1725	961	
664	2	89	304	

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
Sp	ID	(nt)	(nt)
667	1	3	413
668	1	1	330
671	2	612	516
671	1	3	338
674	2	865	584
679	1	1	237
679	3	1569	1906
688	1	1236	835
688	2	1352	1077
694	1	3	143
696	2	818	432
706	1	36	223
709	3	1183	1449
711	1	3	908
715	1	3	167
716	1	2	637
723	1	133	570
723	1	763	383
723	1	1656	829
723	2	1498	1112
727	1	2	472
729	1	268	441
731	1	130	828
735	1	2	214
736	1	3	782
738	1	2	298
742	1	3	230

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins					
CoDing	ORF	Start	Stop		
1b	1D	(nt)	(nt)		
745	3	1148	1780		
746	2	282	1464		
749	1	685	344		
751	1	901	452		
755	1	97	522		
755	2	520	918		
759	2	663	400		
764	2	1033	746		
767	1	1	405		
768	1	1	2	373	
771	1	1	1058	534	
778	1	1	1735	902	
765	1	1	1790	1031	
787	1	1	1250	631	
791	1	1	3	224	
799	1	1	15	1260	
804	1	1	304	711	
805	1	1	3	680	
808	1	1	219	842	
810	1	1	2221	1312	
810	2	1	304	711	
812	1	1	38	979	
817	1	1	714	358	
818	2	1	487	1104	
819	2	1	1529	1032	
819	3	1	1746	1419	
820	1	1	195	1064	

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
ID	ID	(Int)	(Int)
828	1	506	255
829	1	48	800
830	1	578	291
832	1	594	298
835	1	320	796
840	3	491	709
845	1	912	457
850	2	303	449
853	1	715	359
860	1	2	256
864	1	18	410
864	2	183	715
864	6	1676	1828
870	1	1	5AB
873	1	906	454
875	1	584	294
877	1	1661	1020
878	1	981	544
879	1	1567	785
881	1	1	243
882	1	389	604
890	1	2	508
905	1	793	398
906	1	852	544
912	1	373	188
913	1	3	290
913	2	1092	547

TABLE 3

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins					
Contig	ORF	Start	Stop		
ID	ID	(nt)	(nt)		
915	1	6	151		
915	2	169	402		
921	1	126	386		
927	1	1578	808		
928	1	2	385		
929	1	2	400		
932	1	2	400		
934	1	1	384		
936	1	1052	526		
937	1	2	616		
945	1	220	645		
945	2	649	1242		
946	1	1702	950		
949	1	1	270		
951	1	3	362		
955	1	3	143		
960	1	723	400		
963	1	1	162		
965	1	690	346		
966	1	1079	606		
969	1	3	102		
971	1	12	170		
974	1	319	161		
976	1	692	348		
977	1	2	211		
982	1	1926	982		
984	1	589	296		

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - putative coding regions of novel proteins not similar to known proteins

Conseq	ORF	Start	Stop
id	id	(nt)	(nt)
987	1	3	467
993	1	1	525
994	1	920	549
1004	1	557	318
1014	1	624	313
1015	1	2	463
1016	1	288	145
1019	1	1205	660
1022	1	839	474
1024	1	595	299
1024	2	276	431
1030	1	673	338
1032	1	1355	179
1040	1	794	329
1043	1	3	269
1044	2	115	399
1047	1	1	159
1051	1	704	354
1051	2	1233	733
1063	1	2	400
1069	1	2	146
1069	2	769	533
1075	1	707	399
1077	1	97	405
1081	1	58	438
1086	1	1	384
1087	2	246	431

5

10

15

20

25

30

35

40

45

50

55

S. nigerus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF Id	Start ID	Stop (nt)	Stop (nt)
	1088	1	3	174
	1096	1	474	238
	1098	1	1015	509
	1100	1	1020	531
	1100	2	1520	1158
	1101	1	703	353
	1102	1	385	194
	1107	1	2	510
	1114	1	3	422
	1115	1	2	1268
	1119	1	22	267
	1129	1	40	342
	1132	1	360	181
	1133	1	609	376
	1144	1	446	225
	1147	1	558	280
	1153	1	1	153
	1154	1	3	818
	1159	1	1	330
	1161	1	343	186
	1164	1	427	254
	1171	1	19	240
	1171	2	106	299
	1183	1	2	379
	1195	1	355	179
	1196	1	1	189
	1200	1	33	197

TABLE 3

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aurus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start ID	Stop ID	Start (nt)	Stop (nt)
1203	2	129	464		
1222	2	105	401		
1232	1	1	1	387	
1240	1	2	2	175	
1247	1	520	311		
1271	1	412	221		
1286	1	2	595		
1295	1	1	1	165	
1306	1	1	367	185	
1314	12	158	631		
1316	1	58	570		
1359	1	386	193		
1370	1	1	1	402	
1371	1	1	1	345	
1374	1	710	357		
1378	1	1	2	400	
1392	1	3	1	413	
1411	1	202	432		
1433	1	311	167		
1450	1	1	2	256	
1453	1	295	149		
1471	1	721	398		
1477	1	869	639		
1502	1	794	399		
1518	1	126	449		
1534	1	283	143		
1546	1	3	1	401	

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)
1547	1	506	255
1583	1	3	350
1587	1	3	563
1602	2	170	679
1629	1	1	402
1665	1	468	235
1760	1	625	314
1762	1	3	200
1876	2	119	286
1895	1	2	379
1931	1	798	400
1976	2	715	383
2055	2	252	401
2056	1	131	167
2150	1	523	263
2157	1	794	1399
2164	1	564	283
2175	1	210	400
2212	1	492	331
2338	1	732	367
2342	1	3	167
2352	1	330	166
2352	1	622	398
2355	1	47	352
2356	1	679	141
2359	1	304	152
2431	1	296	150

TABLE 3

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
ip	id	(nt)	(nt)
3046	1	367	185
3049	1	553	278
3050	1	3	314
3052	1	504	253
3065	1	2	157
3070	1	157	190
3075	1	440	322
3080	1	1	285
3092	1	320	162
3093	1	411	250
3100	1	52	237
3103	1	47	298
3118	1	344	174
3123	1	2	145
3127	1	1	147
3128	1	316	169
3142	1	388	203
3144	1	664	386
3151	1	337	270
3155	2	202	384
3168	1	12	176
3205	1	268	145
3222	1	1	150
3203	2	239	400
3371	2	211	199
3598	1	2	148
3598	2	36	1401

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop	Size (nt)
1b				
3046	1	367	105	
3049	1	553	278	
3050	1	3	314	
3075	1	440	222	
3080	1	1	265	
3052	1	504	253	
3092	1	350	162	
3093	1	411	250	
3100	1	52	237	
3103	1	1	47	238
3118	1	344	174	
3123	1	2	145	
3127	1	1	147	
3138	1	336	169	
3142	1	388	203	
3144	1	664	386	
3151	1	337	170	
3155	1	2	202	384
3168	1	12	176	
3205	1	286	145	
3282	1	1	150	
3303	2	239	400	
3371	2	211	399	
3358	1	2	148	
3598	2	36	401	

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
lib	ID	(nt)	(nt)
3568	1	751	377
3595	1	757	380
3618	1	2	238
3618	2	130	402
3622	1	86	358
3622	2	664	398
3642	1	876	439
3649	1	781	398
3651	1	625	314
3664	1	467	637
3674	1	55	402
3677	1	619	311
3704	1	1	402
3726	1	515	269
3765	1	510	256
3779	1	554	357
3794	1	266	135
3794	2	667	377
3796	2	638	375
3801	1	474	262
3806	1	453	298
3807	1	42	389
3815	1	798	400
3827	1	3	320
3842	1	781	392
3853	1	671	389
3855	1	1	324

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
	ID	(nt)	(nt)
3857	1	2	235
3861	1	590	297
3865	1	695	1399
3877	1	3	173
3897	2	143	400
3898	2	125	401
3921	2	103	342
3927	1	70	375
3930	1	76	234
3946	2	651	382
3951	2	105	377
3965	1	646	344
3973	1	795	400
3981	1	3	311
3998	1	3	356
4001	1	481	296
4003	1	90	335
4018	1	2	259
4018	2	186	401
4021	1	1	345
4043	1	3	344
4054	1	3	344
4056	1	1	150
4070	1	1	324
4072	2	187	390
4073	1	1	285
4077	2	127	372

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

5
10
15
20
25
30
35
40
4550
55

contig	ORF	start	stop
ID	ID	(nt)	(nt)
4083	1	3	159
4090	1	27	368
4101	1	103	297
4105	1	1	306
4107	1	570	286
4119	1	629	339
4121	1	740	372
4123	1	3	230
4127	1	3	141
4128	1	2	331
4130	1	768	415
4146	1	97	381
4157	1	3	206
4186	1	505	254
4224	1	510	256
4239	1	1	348
4242	1	709	356
4252	1	589	295
4253	1	1	174
4256	1	568	323
4287	1	303	163
4298	2	498	334
4299	1	471	319
4302	1	153	305
4304	1	1	186

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

F _q 9ig	ORF	Start	Stop
Id	[D]	(nt)	(nt)
4304	1	96	314
4306	1	2	151
4318	1	576	289
4322	1	5	148
4321	1	439	221
4323	1	528	364
4328	1	728	399
4346	1	471	277
4357	2	117	311
4373	1	2	268
4381	1	574	326
4384	1	614	309
4397	1	9	311
4402	1	1	249
4403	1	606	328
4406	1	3	317
4411	1	2	280
4411	2	697	398
4412	1	2	364
4418	1	1	230
4424	1	601	398
4443	1	427	215
4471	1	643	323
4478	1	540	271
4482	1	50	289
4489	1	601	302
4491	1	12	206

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF ID	Start (nt)	Stop (nt)
4490	1	120	306
4501	1	493	248
4518	1	1	246
4526	1	480	241
4527	1	2	163
4532	1	3	239
4542	1	11	175
4547	1	36	200
4573	1	1	231
4598	1	642	322
4619	1	1	180
4630	1	349	176
4652	1	1	246
4659	1	2	157
4680	1	28	183
4690	1	144	174

Table 4

5	ORF	SEQ ID NO	BLAST HOMOLOG	Antigenic Regions			
				Region 1	Region 2	Region 3	Region 4
168_6	5192	lipoprotein		36-45	84-103	152-161	176-185
238_1	5193	chrA		21-39	48-58	84-95	232-249
51_2	5194	OppB gene product (B. sub		20-36	70-79	100-112	121-131
278_3	5195	lipoprotein 1		20-29	59-73	85-97	162-171
276_2	5196	lipoprotein		21-33	65-74	177-186	211-220
45_4	5197	ProX		28-37	59-69	85-100	120-129
315_8	5198	hypothetical protein		45-54	88-97	182-192	243-253
154_15	5199	unknown		31-40	48-58	79-88	95-104
228_3	5200	unknown		25-38	40-52	64-74	80-89
228_6	5201	unknown		29-41	89-101	128-143	173-184
50_1	5202	unknown		21-33	52-61	168-182	197-206
112_7	5203	iron-binding periplasmic		21-31	58-67	92-101	111-120
442_1	5204	unknown		30-39	91-100	122-137	182-192
66_2	5205	unknown		50-59	104-116	127-136	167-182
304_2	5206	Q-binding periplasmic		19-28	48-57	75-84	103-116
44_1	5207	hypothetical protein		27-36	86-95	129-138	192-201
161_4	5208	SphX		27-44	149-161	166-175	201-210
46_5	5209	cmpC (permease)		21-33	61-70	83-92	100-109
942_1	5210	trah [Plasmid pSK41]		83-92	109-118	127-142	
5_4	5211	ORF (S. aureus)		12-22	87-96	111-120	151-160
20_4	5212	peptidoglycan hydrolase (S		24-34	129-138	141-150	161-171
328_2	5213	lipoprotein (H. flu)		81-90	123-133	290-299	
520_2	5214	fibronectin binding protein		44-54	63-79	81-90	95-110
771_1	5215	emm1 gene product (S. pyo		30-39	65-82	96-106	112-121
999_1	5216	predicted trithorax prot. (D		7-16	120-129	157-166	
853_1	5217	ORF2136 (Marchantia polyr		43-52	88-97	102-111	
287_1	5218	psaA homolog		13-22	28-44	72-82	114-124
288_2	5219	cell wall enzyme		14-23	89-98		
596_2	5220	penicillin binding protein 2b		40-49	59-68	76-87	106-115
217_5	5221	fibronectin/fibrinogen bindii		28-37	40-49	62-71	93-111
217_6	5222	fibronectin/fibrinogen bp		10-19	31-40	54-62	73-92
528_3	5223	myosin cross reactive prote		4-13	29-47	60-73	90-99
171_11	5224	EF		20-31	91-110		
63_4	5225	penicillin binding protein 2b		12-21	59-68	95-104	
353_2	5226			46-55	62-71		
743_1	5227	29 kDa protein in fimA regi		23-32	68-79	94-103	175-184
342_4	5228	Twitching motility		10-19	48-60	83-92	111-121
69_3	5229	arabinogalactan protein		97-106	132-141	158-167	180-189
70_6	5230	nodulin		36-45	48-57	137-160	179-188
129_2	5231	glycerol diester phosphodie		8-17	41-50	55-74	97-106
58_5	5232	PBP (S. aureus)		26-35	70-79	117-126	152-161
188_3	5233	MHC class II analog (S. aure		72-81	94-103	115-124	136-145
236_6	5234	histidine kinase domain (Dic		24-33	52-67	81-94	106-121
310_8	5235	clumping factor (S. aureus)		59-71	77-86	93-102	118-127
601_1	5236	novel antigen/ORF2 (S. au		45-54	91-104	108-117	186-195
544_3	5237	ORF YJR151c (S. cerevisiae		76-90	101-111	131-140	154-164
662_1	5238	MHC class II analog (S. aure		22-32	71-80	89-98	114-122
87_7	5239	S' nucleotidase precursor ('		29-45	62-71	105-114	125-137
120_1	5240	B65G gene product (B. sub		102-111			

EP 0 786 519 A2

Table 4

5	ORF	Antigenic Regions (cont)				
		Region 5	Region 6	Region 7	Region 8	Region 9
10	168_6	244-272	303-315			
	238_1	260-269	291-301	308-317		
	51_2	140-152	188-208	211-220	256-266	273-283
15	278_3	198-209				
	276_2	255-268				
	45_4	177-199	221-230	234-243	268-279	284-293
	316_8					
	154_15	148-157	177-187	202-211		
20	228_3	101-119	139-154	166-181		
	228_6					
	50_1					
	112_7	136-149	197-211	218-229	253-273	
	442_1	199-210	247-257	264-277	287-309	
25	66_2					
	304_2	178-187	250-259			
	44_1					
	161_4					
	46_5	131-141	162-176	206-215	243-252	264-273
30	942_1					
	5_4	189-205	230-239	246-264	301-318	340-354
	20_4	202-212	217-234	260-275	314-336	366-373
	328_2					
	520_2					
35	771_1	145-154				
	999_1					
	853_1					
	287_1	154-164				
	288_2					
40	596_2	121-130				
	217_5	244-253	259-268	288-297	302-311	
	217_6	144-158	174-183	188-197	207-216	226-242
	528_3					
	171_11					
45	63_4					
	353_2					
	743_1	197-207				
	342_4					
	69_3	195-211				
50	70_6	206-215	263-272	291-301	331-340	358-371
	129_2	117-127	141-157	168-183	202-211	222-231
	58_5	184-203	260-269	275-299	330-344	372-381
	188_3					
	236_6	138-147	163-172	187-198	244-261	268-278
55	310_8	131-140	144-153	177-186	190-199	204-213
	601_1	208-218				
	544_3	170-179	184-193	224-235	274-287	327-336
	662_1					
	87_7					
	120_1					

Table 4

5	ORF	Antigenic Regions (cont)					
		Region 11	Region 12	Region 13	Region 14	Region 15	Region 16
	168_6						
	238_1						
10	51_2						
	278_3						
15	276_2						
	45_4						
	316_8						
	154_15						
20	228_3						
	228_6						
	50_1						
25	112_7						
	442_1						
	66_2						
	304_2						
	44_1						
	161_4						
30	46_5	306-315					
	942_1						
35	5_4	393-407	416-426	456-465			
	20_4	396-405	410-419	461-481			
	328_2						
40	520_2						
	771_1						
	999_1						
	853_1						
	287_1						
45	288_2						
	596_2						
	217_5						
	217_6						
	528_3						
50	171_11						
	63_4						
	353_2						
	743_1						
	342_4						
55	69_3						
	70_6	453-471	506-515				
	129_2	296-315					
	58_5						
	188_3						
	236_6	358-377	410-423	428-439	442-457	467-476	480-493
	310_8	238-251	256-275	281-290	296-310	314-333	338-347
	601_1						
	544_3						
	662_1						
	87_7						
	120_1						

Table 4

5	ORF	Antigenic Regions (cont)				
		Region 17	Region 18	Region 19	Region 20	Region 21
168_6						
238_1						
51_2						
278_3						
276_2						
45_4						
316_8						
154_15						
228_3						
228_6						
50_1						
112_7						
442_1						
66_2						
304_2						
44_1						
161_4						
46_5						
942_1						
5_4						
20_4						
328_2						
520_2						
771_1						
999_1						
853_1						
287_1						
288_2						
596_2						
217_5						
217_6						
528_3						
171_11						
63_4						
353_2						
743_1						
342_4						
69_3						
70_6						
129_2						
58_5						
188_3						
236_6						
310_8	357-366	370-379	429-438	443-452	478-487	551-560
601_1						
544_3						
662_1						
87_7						
120_1						

Table 4

5	ORF	Antigenic Regions (cont)					
		Region 23	Region 24	Region 25	Region 26	Region 27	Region 28
10	168_6						
	238_1						
	51_2						
15	278_3						
	276_2						
	45_4						
	316_8						
	154_15						
20	228_3						
	228_6						
	50_1						
	112_7						
25	442_1						
	66_2						
	304_2						
	44_1						
	161_4						
	46_5						
30	942_1						
	5_1						
	20_4						
	328_2						
35	520_2						
	771_1						
	999_1						
	853_1						
	287_1						
	288_2						
40	596_2						
	217_5						
	217_6						
	528_3						
	171_11						
45	63_4						
	353_2						
	743_1						
	342_4						
	69_3						
50	70_6						
	129_2						
	58_5						
	188_3						
	236_6						
55	310_8	622-632	670-685	708-718	823-836	858-867	877-886
	601_1						
	544_3						
	662_1						
	87_7						
	120_1						

Table 4

	ORF	Antigenic Regions	(cont)
		Region 29	Region 30
5	168_6		
	238_1		
10	51_2		
	278_3		
	276_2		
	45_4		
	316_8		
15	154_15		
	228_3		
	228_6		
	50_1		
	112_7		
20	442_1		
	66_2		
	304_2		
	44_1		
	161_4		
25	46_5		
	942_1		
	5_4		
	20_4		
	328_2		
	520_2		
30	771_1		
	999_1		
	853_1		
	287_1		
35	288_2		
	596_2		
	217_5		
	217_6		
	528_3		
40	171_11		
	63_4		
	353_2		
	743_1		
	342_4		
45	69_3		
	70_6		
	129_2		
	58_5		
	188_3		
50	236_6		
	310_8		
	601_1		
	544_3		
	662_1		
55	87_7		
	120_1		

Table 4

ORF	BLAST HOMOLOG	Antigenic Regions			
		Region 1	Region 2	Region 3	Region 4
46_1	5241 aldehyde dehydrogenase	8-17	36-52	83-96	112-121
63_4	5242 glycerol ester hydrolase (P.)	9-26	57-73	93-107	123-133
174_6	5243 ketopantoate hydroxymethyl	71-80	203-212	242-254	265-274
206_16	5244 ornithine acetyltransferase	1-10	34-43	54-63	194-210
267_1	5245 NaH-antiporter protein (E. t)	120-129	332-347	398-408	
322_1	5246 acriflavin resistance protein	58-75	153-164	203-231	264-284
415_2	5247 transport ATP-binding protein	108-126	218-227	298-308	315-334
214_3	5248 2-nitropropane dioxygenase	123-136	216-233	283-292	297-306
587_3	5249 clumping factor	5-14	43-54	59-68	76-95
685_1	5250 signal peptidase	59-68	72-81	86-95	99-108
54_3	5251 fibronectin binding protein I	23-32	37-46	50-59	89-98
54_4	5252 fibronectin binding protein I	43-52	66-75	95-104	147-156
54_5	5253 fibronectin binding protein I	49-60	81-90		
54_6	5254 fibronectin binding protein I	55-71	82-97	139-158	175-186
328_1	5255 lipoprotein (H. flu)	11-20	61-70	96-105	

Table 4

ORF	Antigenic Regions (cont)				
	Region 5	Region 6	Region 7	Region 8	Region 9
46_1	215-242	333-352	376-385	416-432	471-487
63_4	145-154	191-202	212-223	245-265	274-283
174_6					
206_16	239-259	275-284			
267_1					
322_1	298-319	350-359			
415_2	344-353	371-380	395-404	456-465	486-495
214_3	318-337	365-375			
587_3	106-115	142-151	156-166	173-182	186-198
685_1	113-122	130-145			
54_3	128-138	185-194	217-226	251-260	268-277
54_4	175-188	191-200	203-212	220-229	
54_5					
54_6	220-230	287-304	317-326	344-353	364-373
328_1					378-387

45

50

55

Table 4

5	ORF	Antigenic Regions (cont)					
		Region 11	Region 12	Region 13	Region 14	Region 15	Region 17
	46_1						
	63_4	306-315	319-328	366-376	395-420	453-462	467-476
10	174_6						
	206_16						
	267_1						
	322_1						
	415_2	539-555					
	214_3						
15	587_3	217-226	278-287	318-327	332-342	351-360	377-386
	685_1						
	54_3	316-325	329-345	355-372	387-396	416-425	438-448
	54_4						
	54_5						
20	54_6	396-407	427-436	514-531	541-550	569-578	612-622
	328_1						

Table 4

25	ORF	Antigenic Regions (cont)					
		Region 18	Region 19	Region 20	Region 21	Region 22	Region 23
	46_1						
	63_4	485-500	513-525				
30	174_6						
	206_16						
	267_1						
	322_1						
	415_2						
	214_3						
	587_3	396-405	426-442	459-470	485-494	505-514	531-562
	685_1						
35	54_3	455-462	472-491	517-536			
	54_4						
	54_5						
40	54_6	639-648	673-681	703-715	723-732	749-760	772-788
	328_1						

45

50

55

Table 4

5	ORF	Antigenic Regions (cont)				
		Region 24	Region 25	Region 26	Region 27	Region 28
	46_1					
	63_4					
	174_6					
10	206_16					
	267_1					
	322_1					
	415_2					
	214_3					
15	587_3	567-578	584-601	607-840	844-854	858-870
	685_1					877-886
	54_3					
	54_4					
	54_5					
20	54_6	793-802	811-826	834-848	866-876	893-903
	328_1					907-918

Table 4

25	ORF	Antigenic Regions (cont)	
		Region 30	Region 31
	46_1		
	63_4		
	174_6		
30	206_16		
	267_1		
	322_1		
	415_2		
	214_3		
	587_3	889-911	927-936
	685_1		
35	54_3		
	54_4		
	54_5		
40	54_6	925-944	951-997
	328_1		

45

50

55

SEQUENCE LISTING

5 (1) GENERAL INFORMATION:

10 (i) APPLICANT:

(A) NAME: Human Genome Sciences, Inc.
(B) STREET: 9410 Key West Avenue
(C) CITY: Rockville
(D) STATE: Maryland
(E) COUNTRY: US
15 (F) POSTAL CODE: 20850

20 (ii) TITLE OF INVENTION: *Staphylococcus aureus Poly-nucleotides and Sequences*

25 (iii) NUMBER OF SEQUENCES: 5255

30 (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
(B) COMPUTER: HP Vectra 486/33
(C) OPERATING SYSTEM: MSDOS version 6.2
(D) SOFTWARE: ASCII Text

35 (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:

40 (vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/009,861
(B) FILING DATE: 05-JAN-1996

45 (2) INFORMATION FOR SEQ ID NO:1:

50

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5895 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

10

TCCATTATGA AGTCACAAGT ACTATAAGCT GCGATGTTAC CAATGTTTT TAAAATCCCA	60
GTAATAAAAT CAAAAAAATAA GTTAAATAAT GTATTCAATT TAAGTCCTCC TTAATAAAAGa	120
15 aaataGGTAA TAATGTAATA GCTTCTATTA TGATGCCTAA TTGAATGAAT TGGGCAAATG	180
GCTCTTGAT GATAAGTGTG ATAATGAAAA GGGTTAAACT ACAATAATC GCATAATATT	240
TTTTTCGTTT AATAAGTCGC ACAGGAATGG GCTTCTTTTT AGTTGCTGCA GGAGCATATA	300
20 CTGAGATTAC ACCTAAAGAA ATAACTGTTA AAATAATCAT AATTAAAAAG TTAATATGAA	360
AATTACTAT TACTAAAGGT AAAAGTATAA ATAGTATAAT ACTTTCTACA TAACACCAAA	420
AAGAAGAAGG TGCAATGTGCa CCATGTGCAT GtCTTCTTAT TAAATAAAAT GTTAAATTG	480
25 TAATTAACGT AAACAGAAAA ATGTTAAAAA TATAGGCAAT AGTATACATA ACAATTAATT	540
TACCTATATT TTTAGCTAAG ACCTGCATCC CTAATCGTAC TTGCAAAAT TGAATATGAT	600
30 CTAAGTTATT TCTCTTTGA AGATACTGG CAAACTGGTC AATTCTTTA TCAAAATAAT	660
TCAATTTAC ACCACTCTCC TCACTGTCA TATACGATTT AGTACAATCT TTTATCATT	720
TATTGCCTAA CTGTAGGAAA TAAATACTTA ACTGTTAAAT GTAATTGTA TTTAATATTT	780
35 TAACATAAAA AAATTTACAG TTAAGAATAA AAAACGACTA GTTAAGAAAA ATTGGAAAAT	840
AAATGCTTT AGCATGTTT AATATAACTA GATCACAGAG ATGTGATGGA AAATAGTTGA	900
TGAGFTGTTT AATTTTAAGA ATTTTATCT TAATTAAGGA AGGAGTGATT TCAATGGCAC	960
40 AAGATATCAT TTCAACAATC GGTGACTTAG TAAAATGGAT TATCGACACA GTGAACAAAT	1020
TCACTAAAAA ATAAGATGAA TAATTAATTA CTTTCATTGT AAATTTGTTA TCTTCGTATA	1080
45 GTACTAAAAG TATGAGTTAT TAAGCCATCC CAACTTAATA ACCATGTAAA ATTAGCAAGT	1140
GAGTAACATT TGCTAGTAGA GTTAGTTCC TTGGACTCAG TGCTATGTAT TTTCTTAAT	1200
TATCATTACA GATAATTATT TCTAGCATGT AAGCTATCGT AAACAACATC GATTTATCAT	1260
50 TATTTGATAA ATAAAATTT TTTCTATAATT AATAACATCC CCAAAATAG ATTGAAAAAA	1320
TAACGTAAA ACATTCCTT AATAATAAGT ATGGTCGTGA GCCCCTCCCA AGCTCGCGGC	1380
CTTTTTGTA ATGAAGAAGG GATGAGTTAA TCATCATTAT GAGACCCGCC GTTAAATAT	1440

55

His Lys Ala
1025

5 (2) INFORMATION FOR SEQ ID NO:5255:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 155 amino acids
 (B) TYPE: amino acid
 10 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5255:

20 Gly Glu Lys Cys Met Phe Leu Ala Trp Asn Glu Ile Arg Arg Asn Lys
 1 5 10 15

Leu Lys Phe Gly Leu Ile Ile Gly Val Leu Thr Met Ile Ser Tyr Leu
 20 25 30

25 Leu Phe Leu Leu Ser Gly Leu Ala Asn Gly Leu Ile Asn Met Asn Lys
 35 40 45

30 Glu Gly Ile Asp Lys Trp Gln Ala Asp Ala Ile Val Leu Asn Lys Asp
 50 55 60

Ala Asn Gln Thr Val Gln Gln Ser Val Phe Asn Lys Lys Asp Ile Glu
 65 70 75 80

35 Asn Lys Tyr Lys Lys Gln Ala Thr Leu Lys Gln Thr Gly Glu Ile Val
 85 90 95

Ser Asn Gly His Gln Lys Asp Asn Val Leu Val Phe Gly Val Glu Lys
 100 105 110

40 Ser Ser Phe Leu Val Pro Ser Leu Ile Glu Gly His Lys Ala Thr Lys
 115 120 125

Asp Asn Glu Val Leu Ala Asp Glu Thr Leu Lys Asn Lys Gly Leu Lys
 130 135 140

45 Leu Gly Asp Thr Leu Ser Leu Ser Xaa Xaa Arg
 145 150 155

50

Claims

1. Computer readable medium having recorded thereon a nucleotide sequence of the *Staphylococcus aureus* genome as depicted in SEQ ID NOS:1-5,191, a representative fragment thereof or a nucleotide sequence at least 95 % identical to a nucleotide sequence depicted in SEQ ID NOS:1-5,191.
2. Computer readable medium having recorded thereon any one of the fragments of SEQ ID NOS:1-5,191 depicted in Tables 2 and 3 or a degenerate variant thereof.

3. The computer readable medium of claim 1, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.
4. The computer readable medium of claim 3, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.
5. A computer-based system for identifying fragments of the *Staphylococcus aureus* genome of commercial importance comprising the following elements:
 - 10 (a) a data storage means comprising the nucleotide sequence of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95% identical to a nucleotide sequence of SEQ ID NOS: 1-5,191;
 - (b) search means for comparing a target sequence to the nucleotide sequence of the data storage means of step (a) to identify homologous sequence(s), and
 - 15 (c) retrieval means for obtaining said homologous sequence(s) of step (b).
6. A method for identifying commercially important nucleic acid fragments of the *Staphylococcus aureus* genome comprising the step of comparing a database comprising the nucleotide sequences depicted in SEQ ID NOS: 1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95% identical to a nucleotide sequence of SEQ ID NOS:1-5,191 with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence is not randomly selected.
7. A method for identifying an expression modulating fragment of *Staphylococcus aureus* genome comprising the step of comparing a database comprising the nucleotide sequences depicted in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95% identical to the nucleotide sequence of SEQ ID NOS:1-5,191 with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence comprises sequences known to regulate gene expression.
8. A protein-encoding nucleic acid fragment of the *Staphylococcus aureus* genome,
wherein said fragment comprises the nucleotide sequence of any one of the fragments of SEQ ID NOS: 1-5,191 depicted in Tables 2 and 3 or a nucleotide sequence at least 95% identical to such a nucleotide sequence, or a degenerate variant of any of the aforementioned sequences.
9. The nucleic acid fragment of claim 8 which is DNA.
10. The nucleic acid fragment of claim 8 which is RNA.
11. A vector comprising a fragment of claim 8.
12. A fragment of the *Staphylococcus aureus* genome, wherein said fragment modulates the expression of an operably linked open reading frame, wherein said fragment consists of the nucleotide sequence from about 10 to 200 bases in length which is 5' to any one of the open reading frames depicted in Tables 2 and 3 or a nucleotide sequence at least 95% identical to such a nucleotide sequence or a degenerate variant of any of the aforementioned sequences.
13. A vector comprising a fragment of claim 12.
14. A organism which has been altered to contain any one of the fragments of the *Staphylococcus aureus* genome of claim 8.
15. A method for producing a polypeptide in a host cell comprising the steps of:
 - 55 (a) incubating an organism of claim 14 under conditions where said fragment is expressed to produce said protein, and
 - (b) isolating said protein.
16. An organism which has been altered to contain any one of the fragments of the *Staphylococcus aureus* genome

of claim 12.

5 17. A method for regulating the expression of a nucleic acid molecule comprising the step of covalently attaching to said nucleic acid molecule a nucleic acid molecule consisting of the nucleotide sequence from about 30 to 300 bases 5' to any one of the fragments of the *Staphylococcus aureus* genome depicted in Seq ID Nos:1-5,191 and Tables 2 and 3 or a nucleotide sequence at least 95% identical to such a nucleotide sequence or a degenerate variant of any of the aforementioned sequences.

10 18. A nucleic acid molecule being a homolog of any of the fragments of the *Staphylococcus aureus* genome of SEQ ID NOS:1-5,191 and Tables 2 and 3, wherein said nucleic acid molecule is produced by a process comprising the steps of:

15 (a) screening a genomic DNA library using as a probe a target sequence defined by any of SEQ ID NOS: 1-5,191 and Tables 2 and 3, including fragments thereof;

19 (b) identifying members of said library which contain sequences that hybridize to said target sequence;

20 (c) isolating the nucleic acid molecules from said members identified in step (b).

25 19. A DNA molecule being a homolog of any one of the fragments of the *Staphylococcus aureus* genome of SEQ ID NOS:1-5,191 and Tables 2 and 3, wherein said nucleic acid molecule is produced by a process comprising the steps of:

20 (a) isolating mRNA, DNA, or cDNA produced from an organism;

25 (b) amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from said fragment of said *Staphylococcus aureus* genome to prime said amplification;

30 (c) isolating said amplified sequences produced in step (b).

35 20. A polypeptide encoded by a fragment of claim 8.

30 21. An antibody which selectively binds to any one of the polypeptides of claim 20.

35 22. A kit for analyzing samples for the presence of polynucleotides derived from *Staphylococcus aureus*, comprising at least one polynucleotide containing a nucleotide sequence of any one of the fragments SEQ ID NOS:1-5,191 depicted in Tables 2 and 3 or a nucleotide sequence at least 95% identical thereto or a degenerate variant of any of the aforementioned sequences, that will hybridize to a *staphylococcus aureus* polynucleotide under stringent hybridization conditions, and a suitable container.

40 23. A *Staphylococcus aureus* polypeptide comprising an amino acid sequence identical to an amino acid sequence selected from the group consisting of SEQ ID NOS:5,192 to 5,255 or comprising an amino acid sequence having at least 95% identity to such a sequence.

45 24. A *Staphylococcus aureus* polypeptide antigen comprising at least one epitope derived from a *Staphylococcus aureus* polypeptide selected from the group consisting of SEQ ID NOS:5,192 to 5,255.

50 25. A polypeptide comprising at least one epitope encoded by a *Staphylococcus aureus* amino acid sequence selected from the group consisting of the epitopic sequences listed in Table 4.

55 26. The polypeptide of claim 24 or 26, wherein said polypeptide is fixed to a solid phase.

50 27. A diagnostic kit for detecting *Staphylococcus aureus* infection comprising

55 (a) an isolated polypeptide antigen of claim 24, and

60 (b) means for detecting the binding of an antibody contained in a biological fluid to said antigen.

65 28. A vaccine composition comprising a polypeptide of claim 24 present in a pharmaceutically acceptable carrier.

65 29. A method of vaccinating an individual against *Staphylococcus aureus* infection comprising, administering to an individual the vaccine composition of claim 28.

Figure 1

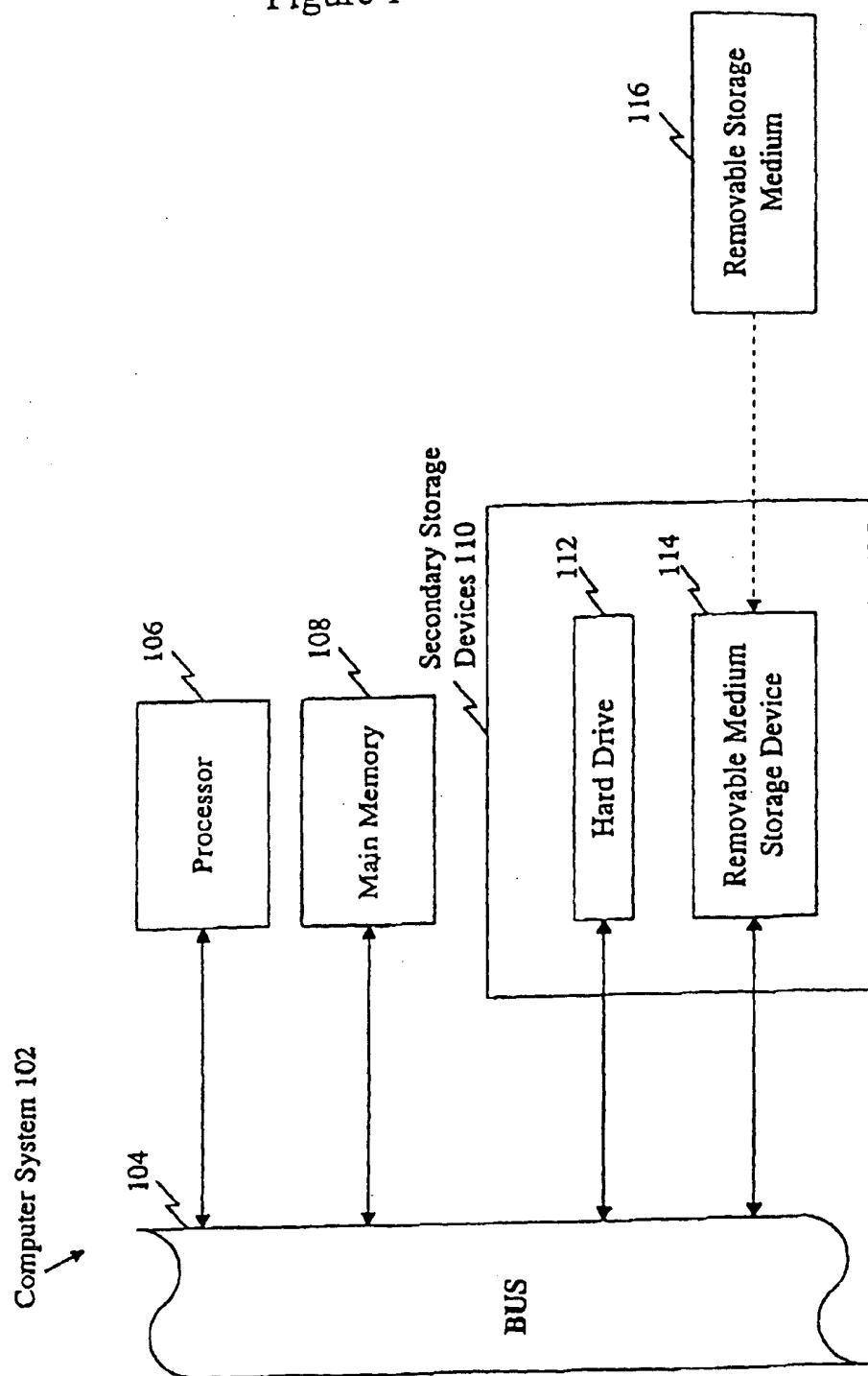


Figure 2

